Functionalization of nucleic acids using an expanded genetic alphabet

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Frank Eggert

aus Erlangen

Bonn 2019

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Michael Famulok

2. Gutachter: Prof. Dr. Olav Schiemann

Tag der Promotion: 15.01.2020

Erscheinungsjahr: 2020

Parts of this thesis have been published in the following journals:

Eggert, F.; Kurscheidt, K.; Hoffmann, E.; Kath-Schorr, S., Towards reverse transcription with an expanded genetic alphabet, *ChemBioChem* **2019**, *20*, 1642–1645 (doi: 10.1002/cbic.201800808).

Domnick, C.; Hagelueken, G.; Eggert, F.; Schiemann, O.; Kath-Schorr, S., Posttranscriptional spin labeling of RNA by tetrazine-based cycloaddition, *Org. Biomol. Chem.* **2019**, *17*, 1805–1808 (doi: 10.1039/C8OB02597E).

Eggert, F.; Kulikov, K.; Domnick, C.; Leifels, P.; Kath-Schorr, S., Illuminated by foreign letters – Strategies for site-specific cyclopropene modification of large functional RNAs via in vitro transcription, *Methods* **2017**, *120*, 17–27 (doi: 10.1016/j.ymeth.2017.04.021).

Eggert, F.; Kath-Schorr, S., A cyclopropene-modified nucleotide for site-specific RNA labeling using genetic alphabet expansion transcription, *Chem. Commun.* **2016**, *52*, 7284–7287 (doi: 10.1039/C6CC02321E).

Domnick, C.; Eggert, F.; Kath-Schorr, S., Site-specific enzymatic introduction of a norbornene modified unnatural base into RNA and application in post-transcriptional labeling, *Chem. Commun.* **2015**, *51*, 8253–8256 (doi: 10.1039/C5CC01765C).

Komisch ist alles, was scheitert.

Bernhard-Viktor von Bülow

Abstract

RNA performs key functions on all levels of life including coding, regulation, and catalysis. These characteristics crucially depend on sensitive three-dimensional structure development. Functional engineering of RNA is a prerequisite to enable non-disruptive probing for the visualization of target sequences and to gather reliable insights into essential features such as their localization and constitution.

This thesis centrally involves genetic code expansion with an unnatural base pair to design novel tools for the site-specific manipulation of functional non-coding RNA.

Particularly, a cyclopropene-equipped unnatural ribonucleoside triphosphate was created that can be applied to efficiently prepare modified RNA via template-directed *in vitro* transcription. Several site-specific labeling approaches were established for the production of functional RNA sequences independent of their length. The cyclopropene moiety serves as highly reactive and minimum-sized handle for ultra-mild post-transcriptional attachment of virtually any reporter group to RNA by means of cell-compatible click chemistry.

Furthermore, distance measurements using electron paramagnetic resonance spectroscopy on highly structured RNA transcripts were facilitated by the introduction of nitroxide spin labels via genetic alphabet expansion transcription. This technique allows investigating structural dynamics of large complexly folded RNAs in nanometer range under solution conditions.

Additionally, significant synthetic progress towards a photoswitchable nucleotide as part of an artificial base pair was achieved. Intriguing applications to spatially and temporally control RNA interactions are envisioned upon completion of this project.

The valuable results presented in this study significantly contribute to the continuous enrichment of knowledge about non-coding RNA molecules that are ubiquitously involved in transmission and expression of genetic information.

Zusammenfassung

RNA erfüllt Schlüsselfunktionen auf allen Ebenen biologischen Lebens einschließlich Kodierung, Regulation und Katalyse. Diese Eigenschaften hängen maßgeblich von fragiler dreidimensionaler Strukturausbildung ab. Das Einbringen chemischer Funktionen ist eine Grundvoraussetzung für störungsfreie Untersuchungen an RNA, um die Visualisierung von Zielsequenzen sowie zuverlässige Einblicke in essentielle Aspekte wie deren Lokalisation und Beschaffenheit zu erlangen.

Diese Arbeit bedient sich der Erweiterung des genetischen Codes mit einem unnatürlichen Basenpaar für die Entwicklung neuartiger Methoden zur positionsspezifischen Manipulation nicht-kodierender RNA.

Insbesondere wurde ein Cyclopropen-funktionalisiertes unnatürliches Ribonukleosid-triphosphat synthetisiert. Dieser Baustein kann effizient in templatgesteuerter *in vitro* Transkription eingesetzt werden, um modifizierte RNA zu erzeugen. Es wurden mehrere positionsspezifische Markierungsstrategien etabliert, die die Produktion funktioneller RNA von beliebiger Länge zugänglich machen.

Durch den hochreaktiven und kleinen Cyclopropenrest können beliebige Reportergruppen post-transkriptional mittels äußerst milder und zellkompatibler Click-Chemie an RNA geknüpft werden.

Des Weiteren wurden Abstandsmessungen in hochstrukturierten RNA-Molekülen durch Transkription mit einem expandierten genetischen Alphabet ermöglicht. Die Einführung von Nitroxid-Spin-Markierungen erlaubte die Anwendung von Elektronenspinresonanz-Spektroskopie. Mit dieser Technik kann die strukturelle Dynamik von großen, komplex gefalteten RNA-Molekülen im Lösungszustand mit Nanometer-Auflösung untersucht werden.

Zusätzlich wurde erheblicher Fortschritt zur Synthese eines photoschaltbaren Nukleotids als Teil eines artifiziellen Basenpaars erzielt. Die Vollendung des Projekts birgt spannende Anwendungen in der zeit- und ortsaufgelösten Steuerung von RNA-Interaktionen.

Die hier präsentierten Ergebnisse tragen erheblich dazu bei, Wissen über nicht-kodierende RNA-Moleküle anzureichern, die zentral in die Transmission und Expression genetischer Informationen involviert sind.

Content

1 Theoretical background

Informational flow in nature is suggested to proceed in one particular direction by the popular ‘central dogma of molecular biology’ (Figure 1).[1] In detail, deoxyribonucleic acid (DNA) serves as storage medium, ultimately holding all the necessary information for the sustenance of life. Ribonucleic acid (RNA) is predominantly known for its intermediary function performed by messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). These species facilitate the biosynthesis of proteins, which are in turn responsible for a vast majority of chemical transformations including DNA replication as well as cellular anabolism and metabolism.

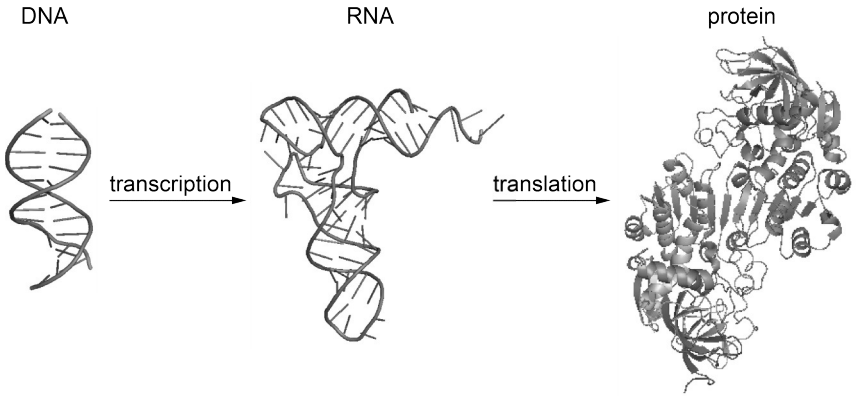


Figure 1. Schematic representation of the biological information flow according to the ‘central dogma of molecular biology’[1]; protein data bank (PDB) files: 1bna[2], 1g59[3], 1hso[4].

From this point of view, DNA poses as the preceding and most important actor. However, ‘*DNA makes RNA makes protein*’, as frequently quoted with regard to the central dogma, has already been revised in several aspects. Looking at the individual properties of the biological trinity, significant additions may be appropriate in order to reflect cutting-edge knowledge. For example, reverse transcription is known as a retroviral feature, but was also found to occur in pro- and eukaryotic life forms.[5–9] Therefore, information can also go in the opposite direction, from RNA to DNA. Furthermore, rRNA demonstrates RNA’s ability to efficiently catalyze chemical reactions. Ribonucleic acid thus performs two essential aspects, namely both information coding *and* catalysis, ensuring the crucial diversity of molecules that is pivotal for the complexity of organisms.[10] To date, a tremendous amount of non-coding RNAs (ncRNAs) has been discovered, which proved to be involved in nearly every cellular pathway with impact ranging from direct chemical catalysis to general cellular signal transduction as well as gene regulation.[11–14]

Based on these extraordinary versatile properties, the so-called RNA world hypothesis evolved, suggesting that life on earth may even originate from an early-world scenario featuring RNA as the key player.

1.1 The RNA world hypothesis

Mankind's intrinsic thirst for the knowledge of its own roots has driven science to identify the first major biopolymer, which once enabled an early self-replicating system and consequently laid the foundation for the evolution of life to occur.

Modern molecular machineries synthesize DNA nucleotides enzymatically by reducing 2′-OH groups of ribonucleic acid building blocks[15,16], demonstrating that there is no DNA without RNA. This led to the generally accepted idea of RNA evolutionary preceding DNA.[14,17–19] Numerous models claim truthful assignment of relevant prebiotic polymers. Among the most popular species are polycyclic aromatic hydrocarbons[20], RNA[10,17,21,22], and a range of artificially developed polymers with modified backbones such as peptide nucleic acid (PNA)[18,23], glycol nucleic acid (GNA)[24], and threofuranosyl nucleic acid (TNA)[25] (Figure 2).

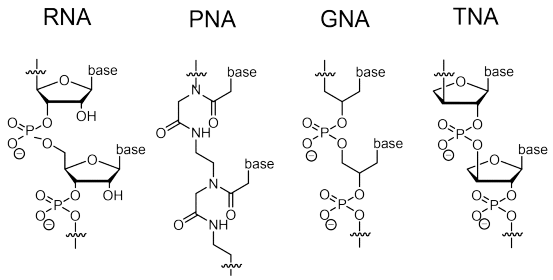


Figure 2. Repetitive units of RNA and artificial biopolymers with potentially prebiotic significance.[19]

While conclusive evidence for any of these theories is hard to deliver, it may also apply that several of the hypotheses are correct and respective molecule species served as an intermediate during a (co)evolutionary process towards the biologically relevant species encountered in the modern world.

It is widely accepted, that an RNA-based era took place prior a DNA-RNA-protein world and this idea has been significantly supported both in a well-thought as well as experimental manner. Miller-Urey-type reaction setups attempted to resemble early-life prebiotic conditions, *i.e.* the composition of available gas molecules, for example CH_4, NH_3, H_2O, CO, and H_2.[26] Mixtures of reasonable *in vitro* atmospheres were subjected to a major energy source such as electric discharge, ultra violet (UV) irradiation, or laser plasma, imitating asteroid impact or volcanic activity.[27–30] Both experiments and *ab initio* calculations resulted in similar molecules formed by this approach. Upon the most prevailing compounds are HCN, formaldehyde, acetylene, and glycine. Starting from these species, formation of all RNA nucleobases via plasma chemistry could be observed and was supported computationally.[27] Although opinions on a correct experimental design are controversial, several attempts demonstrated the faithful emergence of obligate intermediates towards relevant atomic unions.

The more demanding synthesis of ribose scaffolds was also envisioned under prebiotically plausible conditions.[29,31] Polymerization of monomeric compounds such as early nucleoside analogs and inorganic phosphates was suggested to occur non-enzymatically by iterative wet-dry cycles, leading to nucleic acid analogs.[32] Taking together these findings and including the possibility of RNA self-replicating systems[19,30,33–36], it is quite conceivable that RNA-like substances developed billions of years ago (Figure 3).[29] However, it also appears considerable that significant transformations had to be undergone until the complex and unique biopolymer with its distinct Watson-Crick nucleobase pairng patterns was formed in the way we know it.[37]

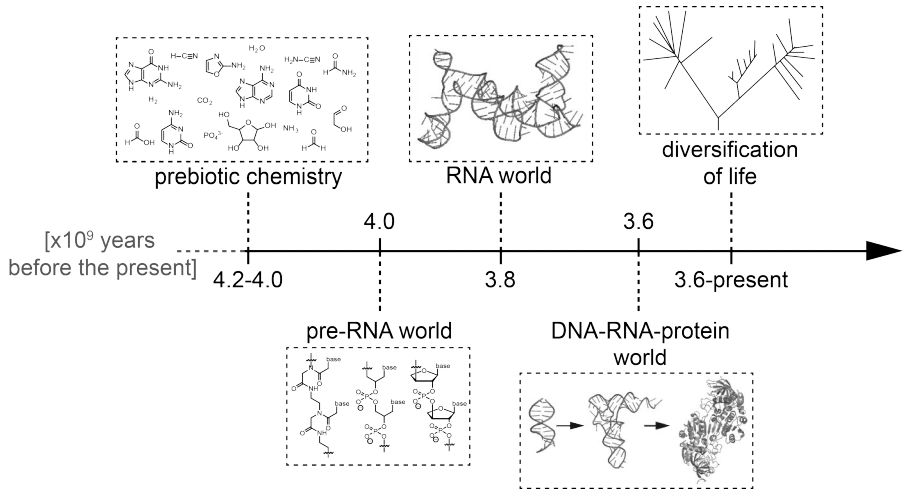


Figure 3. Schematic representation of an early-earth time course scenario embracing the RNA world hypothesis; reproduced from Joyce[19]; PDB files: 2oiu[36], 1bna[2], 1g59[3], 1hso[4].

Several aspects are still contradictory or remain elusive. For example, the high degree of stereo information inherent in β-D-ribose scaffolds is very unlikely to form in a favored way.[37] Further concerns include RNA's fragility under various conditions, which is thought doubtful to endure during harsh environmental influences.[16] However, early-earth conditions such as pressure, temperature, pH etc. can be assumed to have been different from the modern world, possibly contributing to an increased RNA stability. This leads to another bottleneck in the explanation of early-earth RNA existence – the gap between prebiotic chemistry and modern biochemistry.[18] Evidence of primitive archetype peptide enzymes (so-called urzymes) capable of accelerating tRNA aminoacylation along with potential imperfections of an RNA-only historic background further led to the preference of an RNA-protein co-evolutionary focus by many scientists.[14,18,22,38]

Nowadays, RNA is predominantly known for its mediating role between DNA and protein. However, besides protein-coding ribonucleic acids, a plethora of ncRNA species has been discovered in the past decades, which demonstrated that ribonucleic acid is far from being a living fossil.

1.2 Functional non-coding RNA (ncRNA)

RNA has recently emerged as a key player in nearly every cellular pathway.[11,13,14,39] This does not only include the well-known mRNA, tRNA, and rRNA species, but also an ever expanding wealth of other ribonucleic acid molecules with major biological impact. ncRNA is vaguely defined not to participate in protein coding infrastructure. Actually, the majority of transcribed RNA is not involved in the biosynthesis of proteins.[40] However, there is increasing evidence that a considerable share of ncRNA performs important functions and fulfills an extraordinary spectrum of tasks rather than being so-called evolutionary junk.[41–44]

Among several other classes, two types of short ribonucleic acid fragments are known to post-transcriptionally affect gene regulation, namely micro RNAs (miRNAs) and small interfering RNAs (siRNAs). Both are generated by multi-step procession of primary transcripts via similar, but not identical, pathways.[45–47] They comprise small duplex motifs of ~20–30 nucleotides (nt) in length and form complexes with specific proteins that can induce mRNA silencing to repress translation. While siRNAs have been proposed to be involved *e.g.* in antiviral resistance mechanisms[45], miRNAs serve in various endogenous pathways including morphological development and are related to cancer propagation[48,49].

Besides these short ribooligonucleotide species, many classes of ncRNA have been discovered in the past decades and tens of thousands functional ncRNA sequences were identified.[40]

An increasing number of RNA sequences with catalytic activity, so-called ribozymes, has been reported with presence throughout a wide variety of organisms[50] and fuelled the support for an RNA world preceding modern life.[51] Ribozymes form complex tertiary structures with distinct active sites comparable to proteins and are mostly involved in specific nucleic acid scission mediated by general acid-base catalyzed transesterification.[52–54] Modern roles of ribozyme catalysis cannot be overrated, as *e.g.* the well-studied hepatitis delta virus (HDV) self-scissile RNA serves in replication of pathogenic information. Furthermore, gene-regulatory impact via mRNA degradation control has been evidenced in many species including humans.[55] All of these truly fascinating molecules establish their functionality through advanced three-dimensional organization. Exceptional structural motif phenomena such as three-way junctions and pseudoknots are often found as crucial elements. Although more than a dozen of ribozyme classes has been discovered[53], the vast diversity of structural domains unfortunately inhibits predictive knowledge transfer. Thus, investigation of ribozyme architecture has to be carefully addressed in every single case.

Moreover, long non-coding RNAs (lncRNAs) have proven to hold an enormous functional repertoire. Rather set as an arbitrary cut-off to enable differentiation from shorter classes, lncRNAs are defined to contain more than 200 nt and can easily outreach $10 \cdot 10^3$ nt.[42,56] They can be found in almost any sequence context throughout mammalian genomes, *i.e.*

within protein-coding genes, adjacent to them, or far away from translated loci.[40,44,57,58] It has been shown that lncRNAs perform various essential functions. Their scope of regulatory action towards physiological programming includes, but is not limited to, chromosomal dynamics, genome splicing, RNA editing or degradation, and signal transduction.[44] Suppression of certain tumorgenesis pathways in several cancer types has been reported to be mediated by lncRNAs.[56,59,60] Therefore, these species emerged as interesting targets for medical implications.

Despite the apparent abundance and versatility of ncRNAs, precise investigations on particular interaction mechanisms have only been unraveled for few examples. In order to deepen the understanding of this hidden molecular signaling layer, further studies on the localization and structural features that determine essential functional aspects are necessary.

1.2.1 The *glmS* ribozyme

An outstandingly fascinating regulatory RNA is related to the glucosamine-6-phosphate synthase (*glmS*) gene of many bacteria.[61,62] Located in the 5′ untranslated region (UTR) of the corresponding protein-coding mRNA, a small self-cleaving sequence with unique features, known as the *glmS* ribozyme, can be found.

Upon binding of glucosamine-6-phosphate (GlcN6P, **1**) to the RNA, self-cleavage is induced and subsequent degradation of the mRNA is promoted. This results in lower levels of the *glmS* gene, which decreases available GlcN6P (**1**) in a feedback inhibition fashion. GlcN6P (**1**) is a central building block for cell wall synthesis and thus responsible for overall bacterial growth.[63] Since the *glmS* ribozyme contains a small-molecule binding aptamer domain, it simultaneously belongs to the class of riboswitches.[64]

This remarkable non-coding sequence represents the first example of a metabolite-responsive catalytic RNA[61,65] and has been extensively studied in order to reveal its three-dimensional structure as well as its essential requirements to perform cleavage.

To date, hundreds of variants have been described that show major conservation of key structural motifs.[62] A core region of ca. 80 nt is required for scission activity[61,66] (Figure 4A, gray box) and further helices downstream help stabilizing the central domain[67]. Global ribozyme folding and formation of the active site is independent of the cofactor[68], however, catalysis sensitively and specifically depends on the presence of GlcN6P (**1**)[65].

Supported by computational methods[69], the cleavage mechanism involves a guanine nucleobase (G33 in *Bacillus anthracis* and G40 in *Thermoanaerobacter tencongensis*) that acts as proton acceptor in a general acid-base catalysis scheme with GlcN6P (**1**) as proton donor (Figure 4B). This cascade facilitates cleavage between A^{-1} and G^{1}, leaving the separated upstream fragment as 2′,3′-cyclic phosphate (3′-cP).

Crystal structures from different bacterial strains are available[70,71] and exhibit a complex double pseudoknot scaffold (Figure 4C). The *glmS* ribozyme emphasizes versatility and diversity of RNA interaction and reaction scope defined by extensive structure formation. More than being an artifact, catalytic self-cleavage activity has high biological impact and encourages research to implement ribozyme sequences as therapeutic agents.[72]

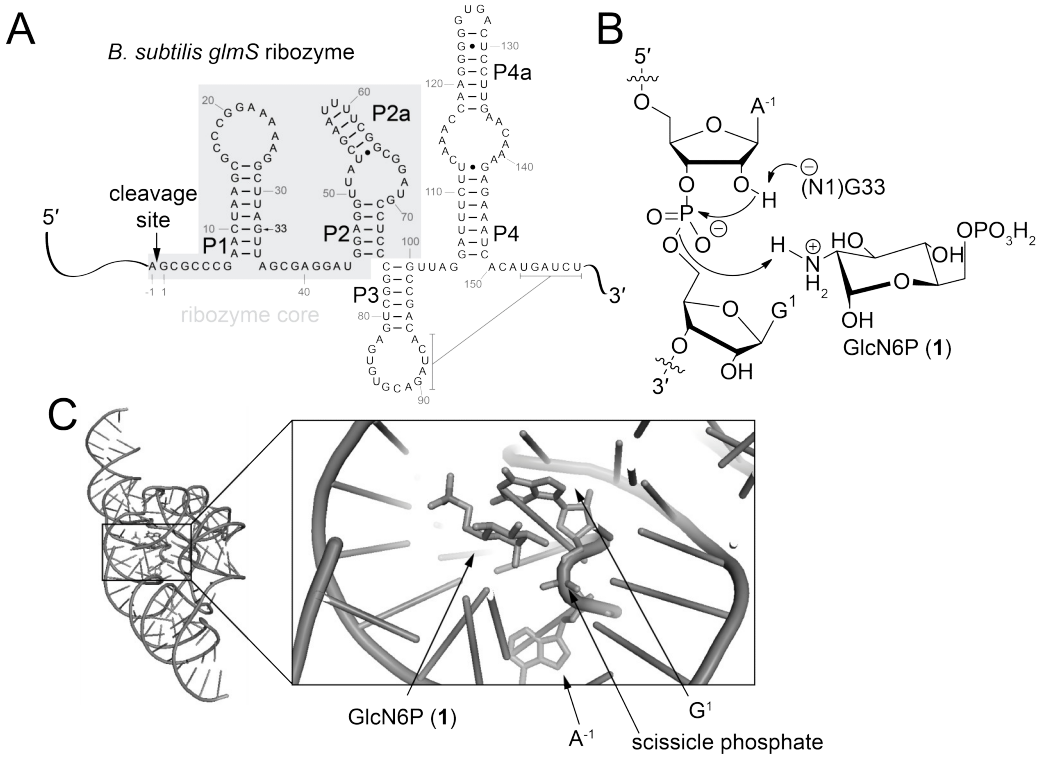


Figure 4. Self-cleavage of the *glmS* ribozyme; A. Secondary structure model of a *B. anthracis glmS* ribozyme, reproduced from Winkler *et al*.[61] B. Schematic representation of a proposed cleavage mechanism[73,74]; C. Crystal structure of a *T. tencongensis glmS* ribozyme bound to its cofactor GlcN6P (**1**) with close-up view of the active site; PDB file: 2z75[71].

1.2.2 The *Xist* lncRNA

X chromosome inactivation is an essential phenomenon in the early development of female mammalian cells.[75–78] From the two existent X chromosomes in each cell, one is transcriptionally silenced to achieve dosage equivalence between males and females. Originating from the *Xist* gene within the X inactivation center of X chromosomes[77], the *Xist* RNA plays a central role in gene silencing. This long ncRNA comprises several thousand kilo bases (kb)[76,77,79] and stands out as an archetype to illustrate the mere limitless possibilities of RNA regulation. Tremendous efforts performed in numerous studies aimed to elucidate the *Xist* RNA's mode of action. Decades of highly laborious approaches indeed yielded a framework of glimpses into the structural features that, rather than the sequence itself, determine its functionality.

So far, the most conserved A repeat region of the *Xist* RNA has been best understood. Several chemical and enzymatic probing endeavors along with computational aid provide information into structural motifs (Figure 5).[79–81] Containing repetitive sequence elements (Figure 5, shaded in gray), the most accurate model of the *Xist* A region reveals a complex map of loops, bulges, helices, and hairpin scaffolds. Despite this advanced mapping along with major contributions that revealed involvement in a multi-step process, still little is known about the distinct mechanism of the *Xist* RNA[82–84]. Playing a crucial role for the X inactivation process, further investigations are likely to continue in order to facilitate comprehension of this fundamental regulatory feature of evolution.

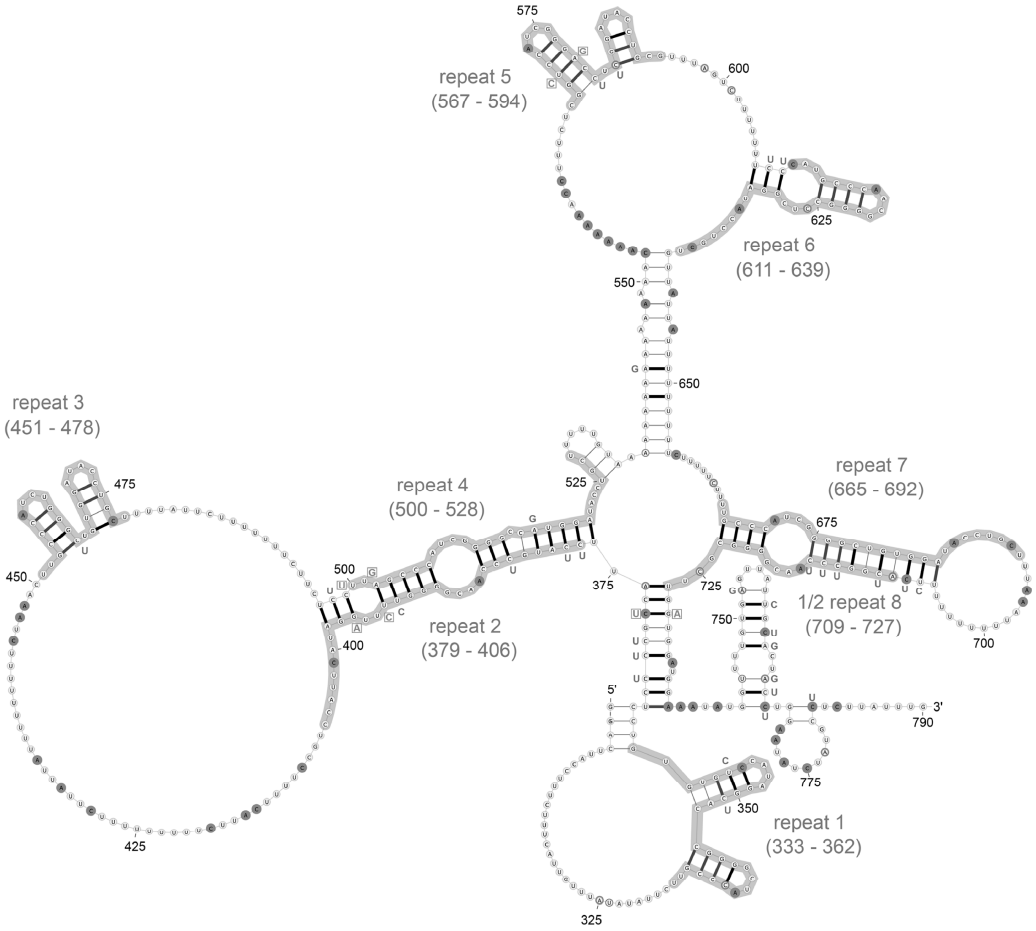


Figure 5. Proposed secondary structure of the Xist A region, taken from Fang *et al*.[81]

These remarkable features of RNA observable at present, whether or not they represent remnants of an early-life scenario, make the unique biopolymer species a fascinating research focus. Advanced comprehension of RNA activities will require knowledge of their structures and conformational dynamics. Investigation of these molecules demands special precision, high efficiency, and low interference with their native performance. Many state-of-the-art techniques are available to gather information on RNA, each of which has special features, advantages, and drawbacks.

1.3 RNA investigation techniques at a glance

In order to visualize or track target RNA sequences and to gain insights into structural features as well as dynamics or modes of action, many sophisticated techniques have been developed. The following section provides an outline of several prominent methods including a brief discussion about their respective scope and limitations. Particular approaches implemented in the experimental part of this thesis are introduced in-depth in section 1.5.

1.3.1 Non-invasive strategies studying RNA structure

RNA molecules, especially functional non-coding RNA, often exhibit complexly folded motifs that are challenging to decipher without disruption. Since unique three-dimensional arrangements define their fascinating characteristics, detailed knowledge about RNA shape is inevitable to understand this species and to gain valuable information on *e.g.* targets susceptive for therapeutic manipulation purposes.

Crystallization

X-ray crystallography is perhaps the most powerful approach acquiring structural data of biomacromolecules with exceptional information density and has centrally driven science ever since its discovery. To date, it is still without alternative receiving high resolution (usually at the lower angstrom range) images of large complexes. Without doubt, this technique has crucially accelerated fundamental research and thus facilitated state-of-the-art understanding of life.

Crystallization holds unreached potential for protein structure solving and also numerous successful investigations on ncRNA such as ribozymes have been reported[36,51,85–87], enormously contributing to the knowledge available. For example, mapping of human immunodeficiency virus (HIV) reverse transcriptase (RT)[88] or the HDV ribozyme[85] (Figure 6) set the structural basis to tackle the life cycle of these life-threatening viruses.

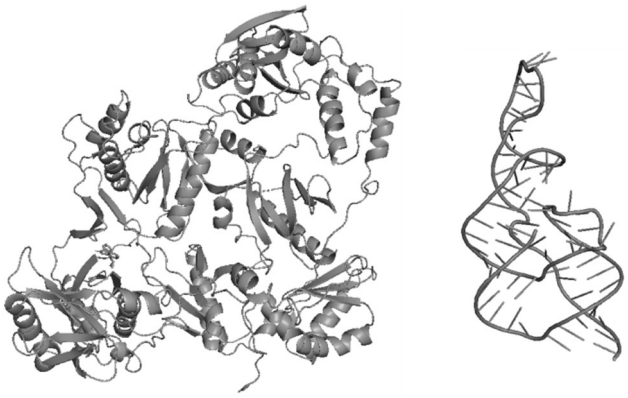


Figure 6. Crystal structures of the HIV RT (left) and the HDV ribozyme (right); PDB files: 3hvt[88], 1drz[85].

However, sample preparation is probably the rate-limiting step investigating RNA. Researchers frequently face problems crystallizing target constructs. Therefore, parts of the construct are often removed, artificially extended, or co-crystallized with proteins to facilitate formation of crystals. *In vitro* transcribed RNAs exhibit considerable sequence and conformation heterogeneity, further interfering with crystal development.[89] While successfully applied to short duplex sequences[90], adaption to large RNAs remains challenging, time-consuming, and may prove inapplicable to certain constructs[87,91]. Arguably, aggregated states may differ from solution conformations due to varying significance of non-covalent interactions in packed crystals.[92] Methods to reliably study large RNA molecules under native (*i.e.* solution state) conditions are thus highly desirable.

Nuclear magnetic resonance (NMR) spectroscopy

Characterization of RNA in solution can be elegantly performed using NMR techniques. Accurate assignment of signals allows for structure determination and has contributed a significant share of data available in the protein data bank as main source for biomacromolecules structures[93]. In combination with computer-aided calculations, precision is achievable in the same range as X-ray crystallography.

Due to hindering overlap of sugar proton signals, mostly amino, imino, and aromatic regions of proton-NMR spectra are used to gain correlation data. Even distances and relative angles between two nuclei are accessible employing multi-dimensional NMR experiments such as nuclear overhauser spectroscopy (NOESY).[94] Although NMR is an important source of information studying nucleic acid sequences, the obtained spectra tend to be of considerable complexity even for small RNAs (Figure 7).

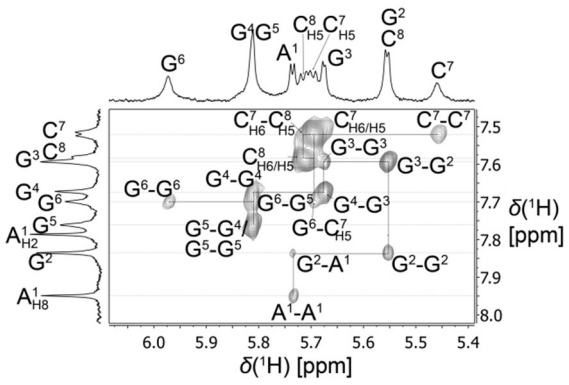


Figure 7. Exemplary two-dimensional NOESY NMR spectrum in the aromatic chemical shift region of an 8-mer RNA G-quadruplex[95].

Partially overcoming these shortcomings, isotopic labeling with ^{13}C and ^{15}N can be applied to enhance facile and selective observations. For short sequences, site-specific labeling of target RNAs is possible via solid-phase synthesis with isotopically doped building blocks.

This approach restricts accessible constructs to the limits of synthetic RNA preparation. Longer constructs are only available by elaborate ligation and purification techniques. *In vitro* transcription with isotopically enriched nucleoside triphosphates can also be used, but will result in multiple labeling positions. Furthermore, relatively large sample amounts (usually milligrams)[96] are required to obtain appropriate signal-to-background ratios, which may not be feasible in a cost-effective way for many targets.

In silico methods

Nowadays, scientific progress is not imaginable without computer-aided three-dimensional modeling of biomacromolecules to complement experimentally-based hypotheses, deepening understanding of wet lab results, and predicting outcomes of potential investigations.

Several approaches are available to gather *in silico* knowledge. In addition to quantum-mechanical methods such as density functional theory (DFT), molecular dynamics (MD) calculations are wide-spread and give most reliable results.[97,98] Using software implementations such as the very fast and ever improving *Groningen machine for chemical simulation* (GROMACS)[99,100] equipped with empirical all-atom parameter sets like the *chemistry at Harvard macromolecular mechanics* (CHARMM)[101] or *assisted model building with energy refinement* (AMBER)[102] force fields, tremendous progress on high quality solution-state simulations of biomolecules has been achieved. Depending on available hardware, modeling of large molecules is potentially possible within few days or less.

RNA computational studies have long been underinvestigated. The intrinsic flexibility of RNA is challenging for MD simulations over a reasonable time span, which is necessary to track folding and relaxation processes as well as to obtain information on stability of the structures. Several distinct energy minima might exist for many constructs and even with different snapshots available, the truth in solution might still be somewhere in between the observed MD data.[103]

Calculations of molecular dynamics are not yet sufficiently mature for *ab initio* studies of large complexes and most methods still need experimental input as basis to optimize minimization of the system's free energy.[98] Experimental support remains essential to prove truthfulness of MD calculations; therefore, the power of simulation applications goes hand in hand with laboratory results. As computational approaches are sometimes not capable of quantitatively reproducing or predicting experimental data, further advances in calculation accuracy will be needed and are expected to continuously drive studies on biomacromolecules.

1.3.2 Labeling and probing approaches

Actively modifying target RNA sequences can efficiently reveal presence, abundance, and localization. Some techniques even allow for structural mapping and elucidation of conformational dynamics.

Radioactive labeling

Probably most-commonly, *in vitro* transcribed RNAs are radioactively labeled with ^{32}P-phosphates for visualization purposes.[48,104] Internal radioactive labeling can be achieved by substitution of one or several ribonucleoside triphosphates (rNTPs) in the transcription mix by [α-^{32}P] rNTPs (Figure 8, upper path). Specific 5′ or 3′ endlabeling of the transcript requires further steps (Figure 8, lower path). As transcripts emerge as 5′-triphosphates, a dephosphorylating step must precede the labeling reaction. Therefore, treatment with alkaline phosphatase is followed by incubation with [γ-^{32}P] adenosine triphosphate (rATP) in the presence of T4 polynucleotide kinase. This leaves the transcript decorated with a single 5′-radioactive phosphate. On the 3′ end, direct attachment of labels is generally possible by *e.g.* unspecific extension with an [α-^{32}P]-bearing dideoxynucleoside triphosphate (ddNTP).

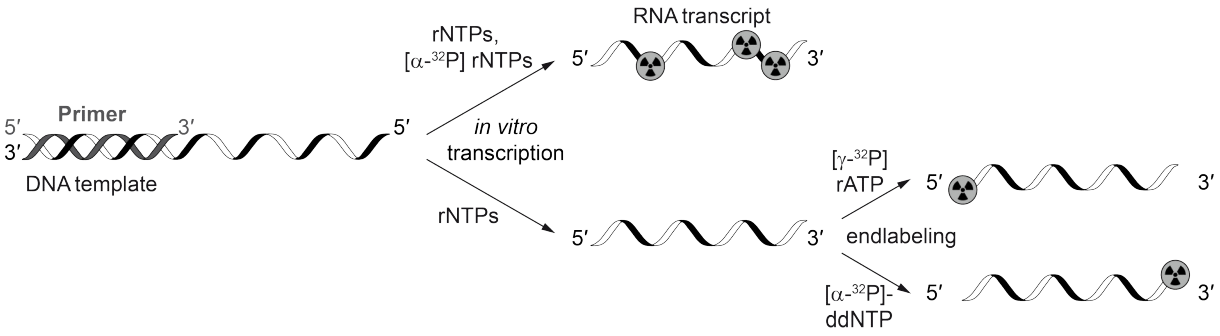


Figure 8. Schematic representation of radioactive RNA labeling approaches.

These widely established approaches benefit from high sensitivity readouts as well as applicability to RNA sequences of any length. Even certain structural information can be gathered. For example investigating self-cleaving ribozymes via radio-endlabeling, fragments reveal whether 5′ or 3′ cleavage occurs.

However, the intrinsic dangers of handling radioisotopes[105,106], restriction of high energy emitters to *in vitro* studies, and significant regulations for in-laboratory use of radioactivity limit the application scope to gather data on RNA.

Fluorescent probing

Structure and localization information about RNA can also be experimentally addressed by several sophisticated fluorescent probing approaches.

Notably, fluorescence *in situ* hybridization (FISH) allows tracking of nucleic acid targets in living cells.[107] For this, short sequences are designed complementary to sections of studied RNAs and *e.g.* labeled with a fluorophore. The probe specifically binds the nucleic acid of interest and thus facilitates detection by fluorescence microscopy. For example, the in-cell localization of *Xist* RNA could be determined to be in the nucleus.[77] However, a major drawback of the method is a high fluorescence background signal.

Advanced techniques make use of Förster resonance energy transfer (FRET). So-called DNA- or RNA-based molecular beacons[108] include a stem region and are equipped with an appropriate fluorophore pair at the 3′/5′ ends (Figure 9). The emission wavelength of one ($\lambda_{1,\,em}$) fluorophore is chosen in the range of the excitation spectrum of the second ($\lambda_{2,\,ex}$). Thus, fluorescence is efficiently quenched when both fluorophores are in close approximation as it is in the closed state (*i.e.* in the absence of a hybridization partner), strongly suppressing background noise. Again, the main sequence is designed complementary to a target RNA so its presence will result in binding to the molecular beacon probe.[109] This dislocates the FRET pair and dramatically increases the fluorescence signal.

There are many modes available, adapted to specific questions on RNA conformational changes, including in-cell applications[110].

Making use of FRET, even RNA dynamics in the low nanometer range can be accessed.[111] Furthermore, several approaches allow for real time observation of systems on single molecule level[112] making it a powerful tool to study various interaction phenomena. Since hybridization probes require careful experimental setup to exclude interference with the native state of sensitively folded RNA structures, FRET pairs are preferentially attached at predefined sites of target molecules via *e.g.* chemical methods for this purpose[113–115].

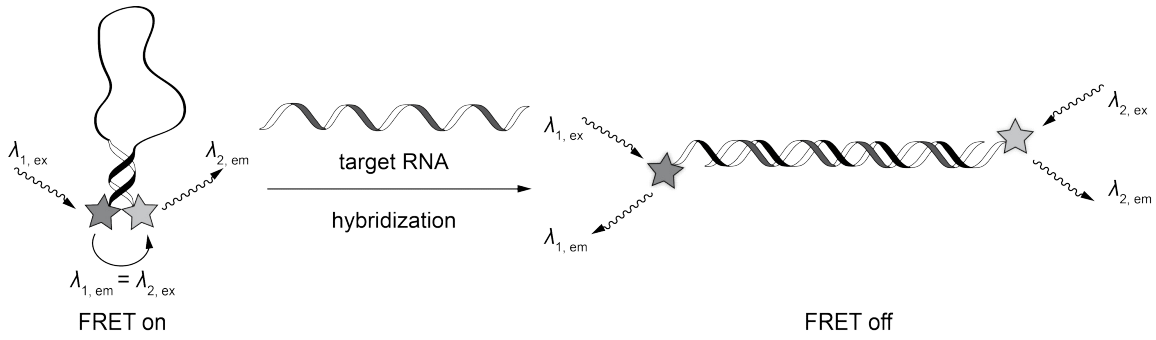


Figure 9. Schematic representation of FRET molecular beacon approach using a hybridization probe.

Chemical and enzymatic probing

Chemical alteration has been widely used to provide insight into structural features of RNAs (Figure 10, upper path). Treatment of target molecules with reactive chemical species can reveal accessible sites thus enabling conclusions on the RNA's secondary structure. For example, exposed nucleobases can be alkylated using dimethyl sulfate (DMS). Employing a set of designed complementary DNA (cDNA) primers, reverse transcription gives information on modification occurrence.[116] Reverse transcriptases (RTs) that are incapable to read over the derivatized sites stall at the alkylated positions. Subsequent denaturing polyacrylamide (DPAGE) analysis visualizes obtained fragments, which can be used for structural mapping of the studied RNA.

In addition to a wide range of chemical probes reacting with nucleobases, the sugar entity, or the phosphate backbone[117,118], standalone enzymatic probing can be applied to give information about secondary structures (Figure 10, middle path). While primer extension assays have gained considerable attention[119], also nucleases are applied to specifically cleave RNA in unpaired regions to fragment the sequence of interest (Figure 10, lower path)[120]. DPAGE separation of the degraded nucleic acid gives information on double stranded vs. single stranded domains, enabling assembly of secondary structure models.

These approaches have tremendously contributed to the knowledge about conformational elements of highly structured RNAs such as the ribosomal units[121]. However, evaluation of data can be laborious and some methods should be carefully interpreted, since *e.g.* cleavage reactions can lead to structural changes that in turn open new sites to be attacked and thus give ambiguous data.[13]

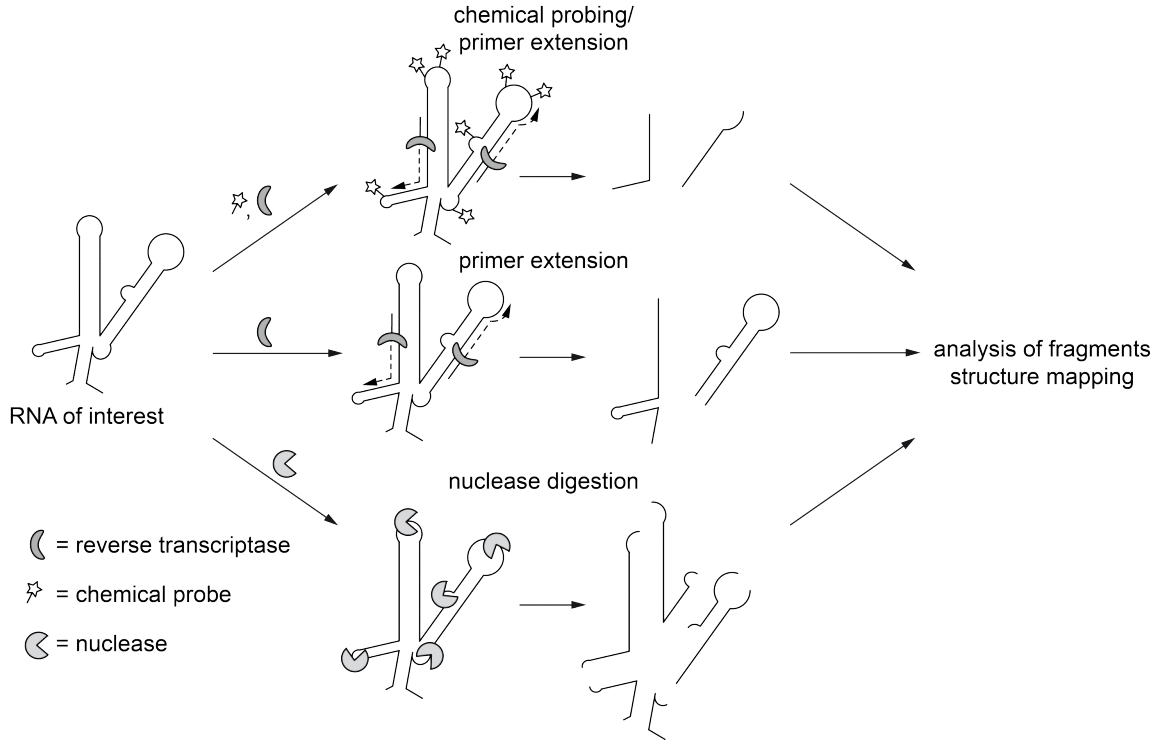


Figure 10. Schematic representation of chemical and enzymatic probing approaches.

Modified nucleotides

Intrinsic and site-specific labeling of target RNAs is extremely useful in order to gather reliable data in a highly specific manner. This can be achieved by the incorporation of modified nucleotides into the sequence to be studied and has been widely applied manipulating RNA molecules.[11]

Chemical solid-phase synthesis of oligonucleotides facilitates quantitative introduction of desired moieties at predefined sites. This technique requires functionalized and protected building blocks. Principally, any scaffold susceptive to RNA solid-phase chemistry could be implemented.[113] To ensure minimal disturbance, however, predominantly 5′-dimethoxytrityl (DMT)-protected 3′-cyanoethyl phosphoramidites (CEPs) of modified canonical nucleosides are employed (Figure 11A)[122]. Unfortunately, not every chemical moiety is stable under the harsh reaction conditions during solid-phase synthesis. Therefore, many approaches involve small handles such as iodine atoms[123,124], alkynes[48,125–127], or azides[126] (*e.g.* Figure 11A, compounds **2**-**6**) that can be post-synthetically converted with high yields to attach the desired reporter groups. Besides delicate preparation and handling of CEPs, chemical synthesis of RNA involves strongly denaturing steps and is currently limited to ca. 120 nt in length (*biomers.net*, requested 2019/08). Site-specific labeling of long non-coding RNAs thus remains inaccessible by this approach. Overcoming this issue, enzymatic ligation has successfully been applied to join several short ribooligonucleotides in order to obtain larger molecules.[122] Unfortunately, rather low coupling yields and laborious purification procedures prevent broad utilization of ligation techniques.

Enzymatic introduction of modifications into the RNA of interest is strongly favorable to allow for preparation of longer constructs. Naturally, just a few small nucleotide modifications are tolerated by RNA polymerases, thus post-synthetic transformations often serve to conjugate sequences with desired reporter groups[128,129]. Most prominently, uridine scaffolds are employed[130], which are enzymatically relatively well-recognized bearing modifications on the 5-position of the nucleobase heterocycle (Figure 11B, compounds **7**-**10**). Using derivatized canonical triphosphates in enzymatic approaches in turn lacks full control of the labeled positions as all naturally occurring sites will be modified.

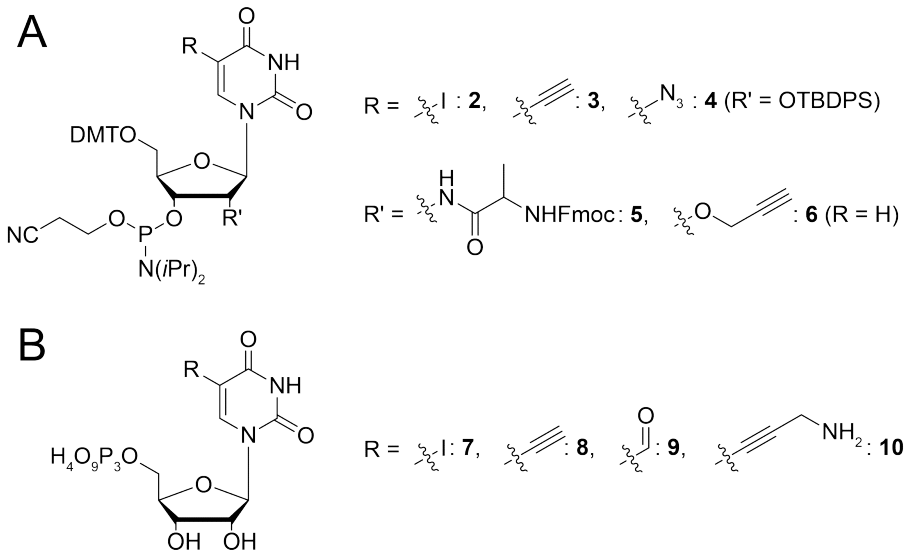


Figure 11. Chemically modified uridine scaffolds; A. Exemplary uridine CEP species for solid-phase RNA synthesis, modified either at the nucleobase (**2**-**4**) or the sugar residue (**5**, **6**)[48,122–127] (TBDPS: *tert*-butyldiphenylsilyl); B. Derivatives of uridine triphosphate for enzymatic incorporation into RNA (**7**-**10**)[128,129,131].

While also enzymatic approaches emerge to post-transcriptionally modify RNAs of interest *e.g.* via alkylation[132,133], all of these methods have specific drawbacks and require careful fine-tuning to successfully manipulate the sequence of interest[113].

Towards the ultimate goal of an efficient site-specific RNA labeling technique that is applicable to arbitrary sequence lengths and offers free choice of reporter groups, an extended information contingent stored in nucleic acids could resolve the current restrictions. Discovery of one or several additional artificial base pairs, which are processed by standard enzymes with similar precision as natural building blocks, potentially even tolerating bio-orthogonally reactive handles, would have enormous potential and set milestones in all bioengineering-related interests.

1.4 Unnatural base pairs (UBPs)

Naturally, scientists seek for possibilities beyond the as-is state. The condensed chemical diversity available in the only four naturally occurring nucleobases virtually demands for expansion investigations. Development of a third or even more artificial base pairs that function in standard polymerase chain reaction (PCR) holds monumental potential. Equally fascinating, the outcome could support the idea that a greater variety of nucleotides once existed and/or make progress toward the goal of (semi-)synthetic life forms. It has nonetheless just been about 30 years ago that serious attempts toward augmenting the natural genetic code started to emerge.

In the late 1980s, Steven Benner and co-workers began to investigate this topic relying on the most obvious choice for such a venture. With the design of the isocytosine:isoguanosine (***iso*-C**:***iso*-G**) base pair they pursued adapting natural Watson-Crick-like systems in favor of an independently functioning, third base pair (Figure 12).[134,135] This isomeric UBP was set out to resemble canonical hydrogen-bonding (H-bonding) patterns in an orthogonal fashion, facilitating proper recognition and procession by highly specific cellular machineries. Predominantly, gathering information on whether the ***iso*-C**:***iso*-G** UBP was suitable to coexist in an early life genetic alphabet was aspired, contributing to investigations on RNA's catalytic potential and thus helping to support the RNA world hypothesis. The Benner lab could show that the Klenow fragment of *Escherichia coli* DNA polymerase I, T7 RNA polymerase, and avian myeloblastosis virus (AMV) reverse transcriptase are able to distinctively accept the artificial base pair. Significant drawbacks were, however, misincorporation of thymine/uracil (**T**/**U**) opposite ***iso*-G** and vice versa as well as considerable deamination of ***iso*-C**.[136]

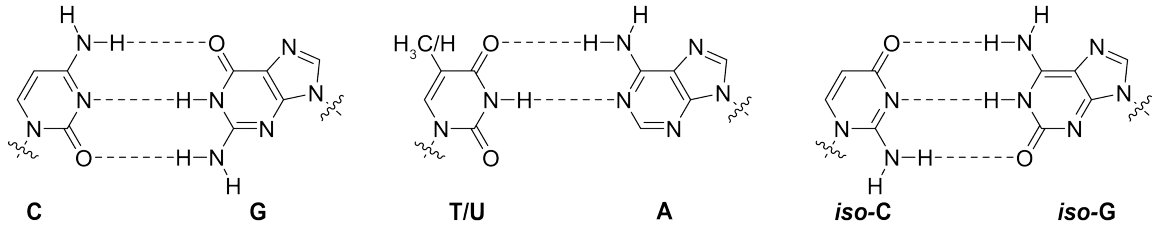


Figure 12. Canonical H-bonding base pairs in comparison with Benner's ***iso*-C**:***iso*-G** UBP; sugar residues are omitted for clarity.

About a decade later, efforts by Juan Morales and Eric Kool revealed that high replication fidelity of base pairs does not necessarily depend on complementary H-bonding patterns.[137,138] Presenting efficient recognition of the non-polar **F**:**Z** system (Figure 13A) by *E. coli*'s DNA polymerase Klenow fragment, these isosteric mimics of thymine (**T**) and adenine (**A**) impressively demonstrated that Watson-Crick type interactions are not an absolute requirement for effective base pairing. Notably, lower selectivity as compared to **T**:**A** was observed, the pyrimidine-shaped **F** nucleobase analog exhibited self-pairing issues, and deoxythymidine triphosphate (d**T** TP) could be incorporated opposite **F** in the template strand.

Nevertheless, researchers all over the world were inspired by this work and soon after Kool's discovery the group of Floyd Romesberg reported a hydrophobic isocarbostyril-derived unnatural base (later entitled **PICS**), that was designed to self-pair in an orthogonal manner to native nucleobases.[139] It stood out as the first stable UBP and DNA duplexes containing a **PICS**:**PICS** pair even exhibited increased melting temperatures compared to canonical

sequences.[140] In order to approach natural-like efficiency and fidelity of UBP replication, however, poor extension rates had to be overcome.

Meanwhile, Ichiro Hirao and coworkers published their first attempt to develop an UBP based on alternative hydrogen-bonding pattern along with steric matching and elimination of interactions with natural entities by attaching bulky groups.

Since all of the first generation unnatural base pairs lacked satisfactory properties in terms of specificity and/or efficiency when processed by natural enzymes, a long and laborious period of iterative screening and optimization procedures was initiated that has now culminated in several state-of-the-art UBPs independently developed by the Benner, Hirao, and Romesberg groups that will be discussed below (Figure 13B).

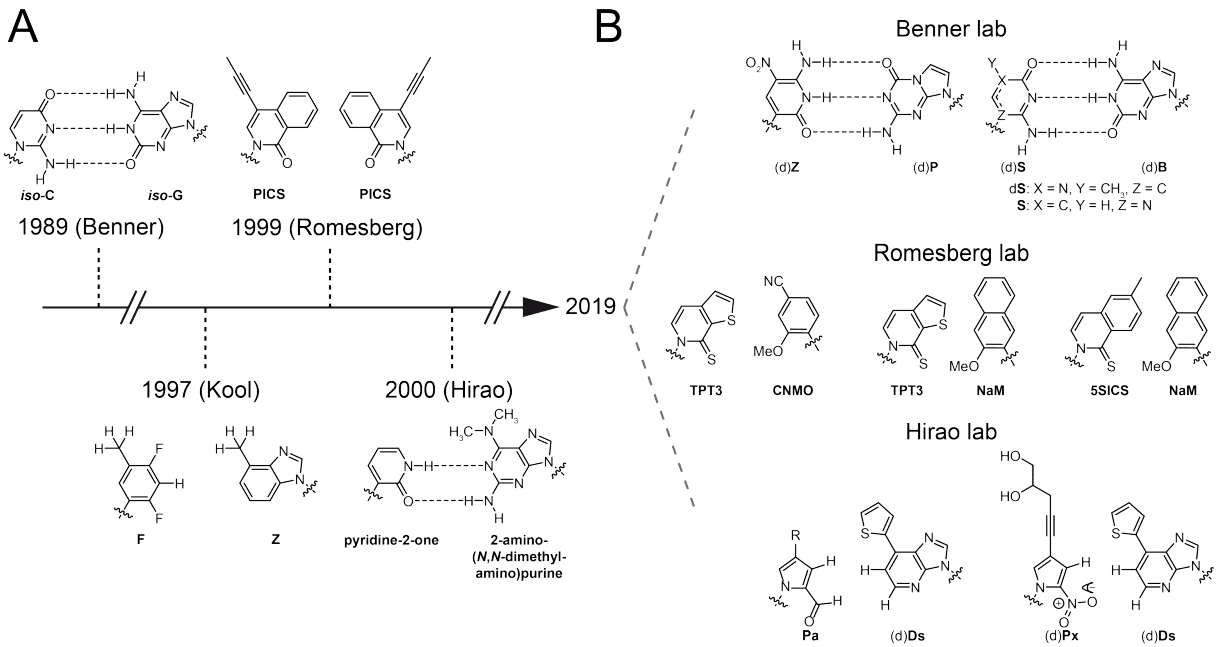


Figure 13. The (condensed) history of UBPs; A. First-generation UBPs; B. Current state-of-the-art UBPs; nucleic acid backbones are omitted for clarity.

1.4.1 Benner's hydrogen-bonding UBPs

The Benner group continued to investigate the ***iso*-C**:***iso*-G** base pair and began to carefully optimize the nucleobases in order to overcome initial shortcomings. The main drawback was thought related to the tautomerism of non-natural nucleobase ***iso*-G**. In the minor conformational state (~10 %)[141], *i.e.* the enol form, ***iso*-G** offers an H-bonding donor-acceptor pattern that develops inconvenient pairing with **T**/**U** rather than with ***iso*-C**. This keto-enol equilibrium seemed to significantly hamper correct polymerase discrimination between three base pairs. Therefore, a 7-deaza variant (**C7*iso*-G**) was synthesized to fix this issue in favor of the non-aromatic keto-form.[141,142]

Another key issue was epimerization of ***iso*-G**, which has been successfully suppressed by adding a nitro group at position 3 of the nucleobase ring system.[142–144]

Further improvements[145,146] have recently led to a genetic alphabet expanded by four unnatural nucleotides establishing a 'Hachimoji' ('eight building blocks') DNA and RNA system including the deoxyribose (d) or ribose (r) units of the **Z**:**P** and the **S**:**B** artificial base pairs (Figure 14).

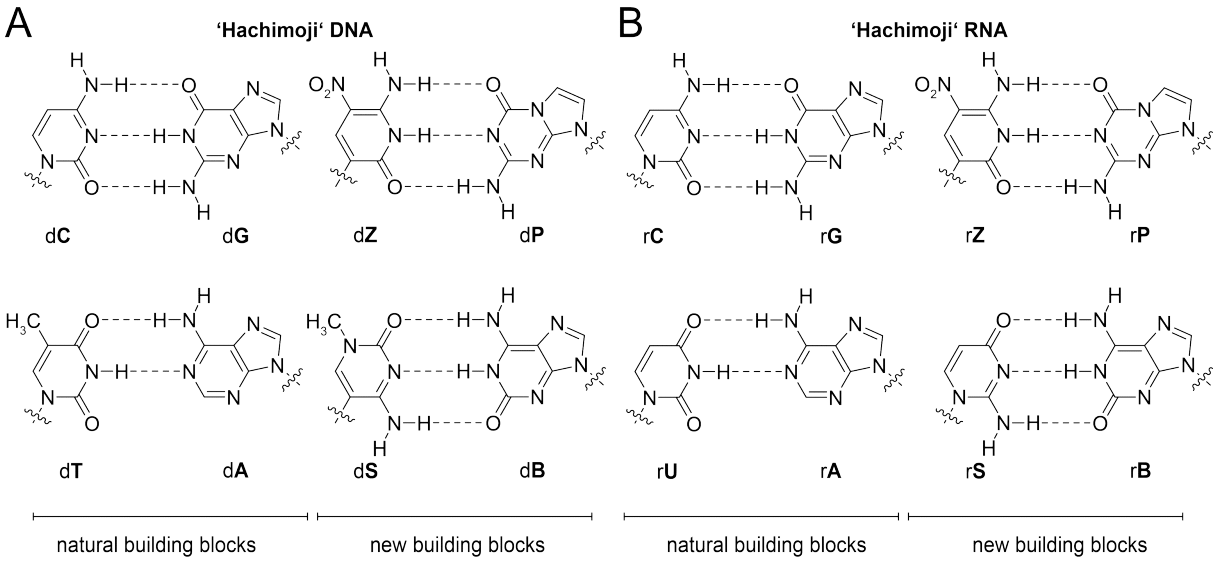


Figure 14. 'Hachimoji' nucleobases; A. Extended DNA alphabet; B. Extended RNA alphabet; reproduced from Hoshika *et al.*[147]; sugar residues are omitted for clarity.

Notably, preparation of respective building blocks to chemically or enzymatically incorporate **Z**,**P**,**S**, and **B** nucleotides into nucleic acids unfortunately require elaborate multi-step syntheses.[136,145,146,148] Hydrogen binding donors (*i.e.* particularly amino groups) demand protective strategies throughout the synthetic route and add additional efforts.

While efficient PCR replication with the eight-letter system has not been demonstrated so far[147], particularly d**Z**:d**P** UBPs have been remarkably implicated into various DNA-based approaches.

Benner and coworkers developed a so-called artificially expanded genetic information system (AEGIS), which is well-analyzed. For example, it was shown that *Thermus aquaticus* (*Taq*) DNA polymerase is able to incorporate up to four consecutive unnatural nucleotides belonging to the d**Z**:d**P** UBP with mutation rates as low as 0.2 % per amplification cycle.[144] Assessing fidelity of incorporation and mutation events, sophisticated digestion and sequencing procedures were developed. The reported values should be sufficient to faithfully replicate d**Z**:d**P** even in plasmid DNA molecules of random context.

Moreover, d**Z**:d**P**-modified sequence libraries have been successfully applied to *in vitro* selection of DNA aptamers (systematic evolution of ligands by exponential enrichment, SELEX) to generate protein-biding sequences with low nanomolar K_D values for *e.g.* an

anthrax-related receptor[149] and a breast cancer cell line[150]. Moreover, transcription and reverse transcription have been demonstrated, paving the way for high-affinity RNA SELEX approaches.[151] This would further broaden the addressable scope of medically relevant targets and add to already extremely useful commercial health care applications such as diagnosis of infections by human immunodeficiency and hepatitis viruses[152].

Supported by several DNA crystal structures[142,147,153,154] (Figure 15), Benner's unnatural base pairs d**Z**:d**P** and d**S**:d**B** are without a doubt truly resembling nature in terms of geometry, pairing efficiency, and thus enzymology. The rather philosophically driven research of Benner and coworkers now resulted in a highly promising system that will allow for several future implementations such as medical research as well as synthetic biology investigations on the origin and further, potentially science-promoted, evolution of life.

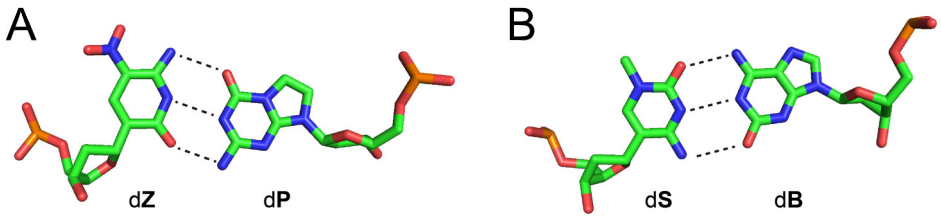


Figure 15. Crystal structure geometries of Benner's Hachimoji deoxyribonucleotide building blocks in a DNA duplex (PDB file: 6mig)[147]; A. UBP d**Z**:d**P**; B. UBP d**S**:d**B** (both UBPs were adjacent in the same 16-mer helical structure and crystallized as host-guest complex with moloney murine leukemia virus (MMLV) reverse transcriptase (RT)).

Despite their high potential serving as artificial expansion for genetic systems, AEGIS components have not yet been reported for in-cell applications. Potentially, corresponding experiments are under investigation[142], as *e.g.* enzymatic phosphorylation of d**Z** and d**P** components has been demonstrated.[155,156] AEGIS investigations in a cellular context would have major impact on the generation of enhanced organisms.

Since Benner's research focuses on fundamental questions such as UBP recognition and procession motifs for in-depth understanding of biochemistry, no implementations were made to couple the artificial nucleotides with reporter groups. Therefore, this particular system is not readily suitable for gathering functional or structural data on ribozymes, for example.

1.4.2 Hirao's UBPs

With the first unnatural base pair by the Hirao group between pyridine-2-one and 2-amino(*N*,*N*-dimethylamino)purine also a hydrogen bonding pattern was created whilst resembling shapes of natural nucleobase geometries.[157] Two nitrogen-bound methyl groups intended to avoid unfortunate cross binding properties with native thymine nucleobases. However, optimization was necessary to truly achieve this goal. Therefore, the artificial analog was soon improved by the introduction of a bulkier thienyl substituent to the purine-like scaffold, inhibiting pairing with canonical nucleobases.[158] Using this enhanced UBP version, later termed **s**:**y**, great progress was achieved, even though insufficient PCR replication properties were reported.[158] The Hirao group was able to site-specifically incorporate r**y** nucleotides into RNA opposite coding d**s** in template DNA via *in vitro* transcription using T7 RNA polymerase. Further, successful cell-free *E. coli* translation could be demonstrated employing an r**s**:r**y**-modified tRNA:mRNA pair with the unnatural nucleotides positioned in the anticodon and codon, respectively, in combination with an artificially charged tRNA bearing an unnatural amino acid.[159] Moreover, selective biotinylation and fluorophore labeling of RNA could be shown by template-directed T7 transcription using a series of functionalized r**y**-substrates.[160–162]

Further efforts aiming towards enhanced possibilities to site-specifically decorate RNA with chemically useful moieties via *in vitro* transcription led to the development of the second-generation hydrophobic UBP **Ds**:**Pa** based on shape complementarity (Figure 16A)[163]. In addition to their packing interactions that enable orthogonal base pairing, H-bonding acceptors were introduced at the rear face of nucleobase scaffolds to allow for interactions with enzymes. Specific and high fidelity PCR replication in DNA was demonstrated, which was generally assessed by means of stalling events during standard Sanger sequencing.[163–165] Efficient incorporation (≥90 %) of chemically modified **Pa** variants (biotin, fluorophores[166]) into RNA showed the high potential of this UBP in labeling approaches for biophysical and cell biology studies.[167]

Notably, **s** and **Ds** nucleobases exhibit intrinsic fluorescence; therefore, successful incorporation into nucleic acids could be directly tracked by fluorescent readout and an advanced d**Ds**:d**Pa** version could even serve as fluorophore-quencher system in a molecular beacon device.[164]

Several **Pa** nucleobase analogs were synthesized and tested for their strong and weak points regarding properties in PCR amplification and *in vitro* transcription.[164] Notably, all of the tested combinations were best recognized and processed by Deep Vent DNA polymerase rather than standard *Taq* DNA polymerase.[168,169] The optimal UBP candidate for DNA replication was found to be **Ds**:**Px** (Figure 16B); **Ds**:**Pa** and pairs with functionalized variants of **Pa** offered superior performance in T7 transcription reactions.[170] For example, a

propynyl-modified **Pa** derivative was efficiently used for post-transcriptional attachment of reporter groups to unnatural base-modified RNAs via copper-mediated click chemistry.[171]

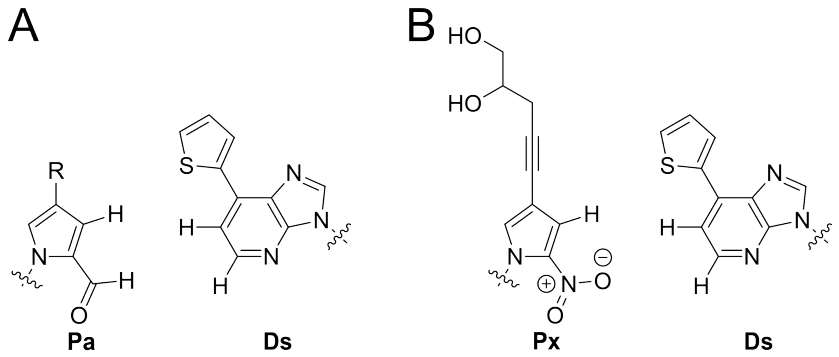


Figure 16. Hirao's second-generation UBPs; A: The **Pa**:**Ds** unnatural base pair; B. The unnatural base pair **Px**:**Ds**; sugar residues are omitted for clarity.

From a synthetic perspective, **Ds** and **Pa** building blocks suitable for chemical or enzymatic access to unnatural base-modified nucleic acids can be prepared in a straight-forward manner (Figure 17).[163] On the one hand, The **Ds** nucleobase analog **11** is available by transition metal-catalyzed addition of a thiophene species to 4-chloro-3-nitro-2-pyridinamine (**12**). Reduction of product **13** to diamine **14** followed by formyl cyclization provides the purine-like scaffold **11**, which can be attached to protected (deoxy)ribose species and thus be converted into either triphosphate (TP) or cyanoethyl phosphoramidite (CEP) derivatives, depending on the respective purpose. On the other hand, **Pa** and derivatives thereof can be synthesized from formyl pyrroles such as **15**, **16** (propyne-substituted), or **17** (iodinated). The latter can easily be decorated with desired residues under Sonogashira conditions. Direct nucleosidation and subsequent standard strategies for TP or CEP buildup facilitate an ample set of functionalized cognates for **Ds**.

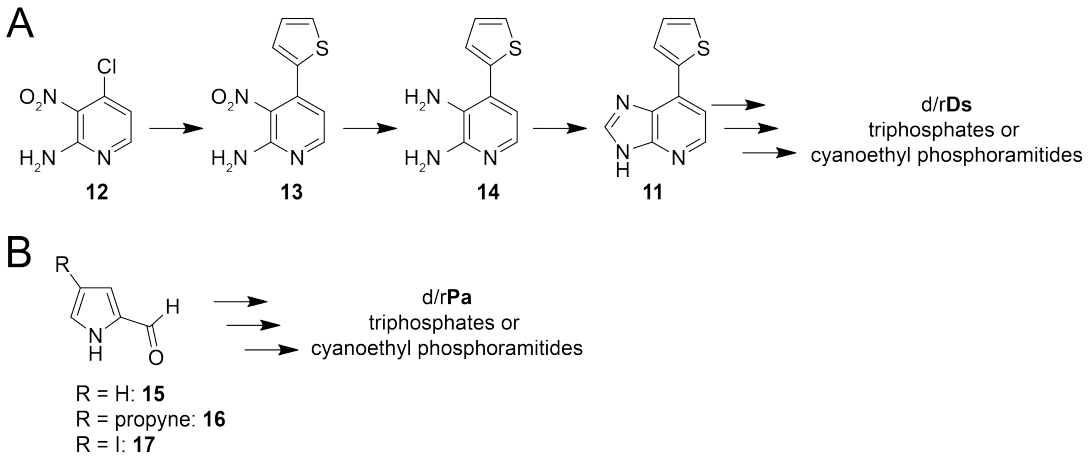


Figure 17. Synthetic approaches towards **Ds** and **Pa** building blocks for solid-phase oligonucleotide preparation or enzymatic incorporation; A. Schematic access to **Ds** derivatives; B. **Pa** and variants can directly be accessed by coupling to (deoxy)riboses followed by TP or CEP synthesis.[163]

With fully functioning UBPs in hand, the Hirao group set out to apply the expanded genetic alphabet performing DNA SELEX (ExSELEX). By increasing the chemical interaction space, facile generation of high affinity aptamers was expected. Indeed, low to sub-picomolar K_D values towards vascular endothelial cell growth factors (VEGFs) were obtained.[172] More than 100-fold increased affinities compared to native sequences could be reached with d**Ds**-modified DNA, verifying the outstanding potential of ExSELEX. Fine-tuning of this approach was achieved introducing a mini-hairpin DNA fragment between the 5′/3′-termini of the aptamers, which stabilized the molecules against nuclease degradation to survive in human serum[173]. A range of cost-effective alternatives to Macugen (an aptamer-based medicine against age-related macular degeneration) independent of Ca^{2+} addition and exhibiting increased thermal and nuclease stabilities as well as enhanced affinities could be evolved with this technique (Figure 18).[174]

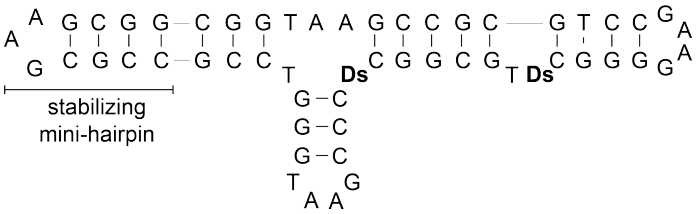


Figure 18. Secondary structure of an optimized anti-VEGF aptamer by Hirao and coworkers.[174]

Versatility of ExSELEX for medicinal applications was further demonstrated by reporting specifically binding aptamer sequences to address several cancer cell lines.[175]

Interestingly, the **Px**/**Pa**:**Ds** UBP systems have not yet been reported to be compatible with cellular machineries such as stable propagation of modified plasmids or transcription and/or translation in live cells. This might be due to the fact that Deep Vent DNA polymerase is uniquely able to specifically replicate d**Px**:d**Ds**-containing sequences. Possibly, further enzyme engineering is necessary to achieve the goal of cell compatibility.

Successful crystallization of the d**Px**:d**Ds** unnatural base pair in Klen*Taq* DNA polymerase was achieved showing nature-resembling edge-to-edge geometry (Figure 19).[176] Hirao and coworkers thus set the structural basis for further biochemical studies employing an augmented genomic alphabet repertoire with a predominantly hydrophobic UBP. These results promise enormous improvements regarding process on therapeutics and diagnostics of challenging threats like cancer and HIV infections.

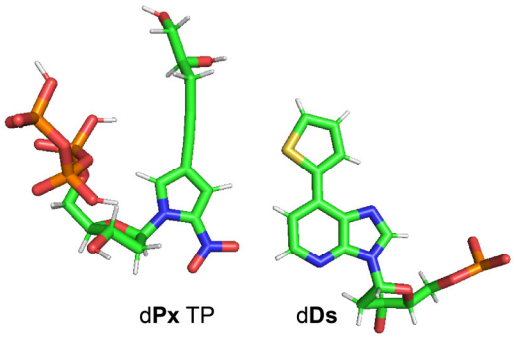


Figure 19. The d**Px**:d**Ds** UBP by Hirao as crystallized using Klen*Taq* DNA polymerase with d**Ds** in a template DNA and an incoming d**Px** triphosphate (PDB file: 5nkl)[176].

1.4.3 Romesberg's hydrophobic UBPs

Unnatural base pair research by the group of Floyd Romesberg has caused international attention. From the beginning, their efforts exclusively relied upon hydrophobic interactions between the nucleobase analogs. With focus on Kool's findings[137,138], hydrophobicity was rated as force strong and selective enough for efficient replication of a UBP system. Initially, Romesberg's unnatural bases were designed only with respect to a certain shape resemblance of native systems reaching the aspired orthogonality. The lack of many multivalent nitrogen and oxygen atoms should overcome tautomerism issues as experienced by the Benner group.[136,139,141,177]

Early experiments involved the **PICS** self-pair (see above) as well as a series of analogs and advanced unnatural nucleobases. DNA replication of these predominantly self-pairing systems was investigated using the Klenow fragment of *E. coli* DNA polymerase, which could recognize and incorporate most of the tested triphosphate substrates, however was unable to extend primer sequences beyond the UBP.[139,140,178,179] As an intermediate achievement, a DNA polymerase mixture with mammalian polymerase β partly helped resolving the extension problem.[180] Based on molecular modeling of a d**PICS** self-pair within a 12-mer DNA duplex, stacked nucleobase analogs were expected rather than native edge-on geometry.[181]

In further studies, a screening of 3600 UBP combinations led to several tens of promising candidates[182–184], which upon structure activity relationship (SAR) evaluation, optimization, and modification revealed **5SICS**: **MMO2** and **5SICS**:**NaM** (Figure 20).

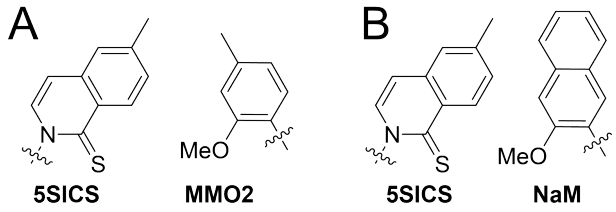


Figure 20. Second-generation UBPs by the Romesberg group; A. **5SICS**:**MMO2**; B. **5SICS**:**NaM**; sugar residues are omitted for clarity.

These scaffolds were PCR-amplifiable in a sequence-independent manner with natural-like efficiency and fidelity of greater than 99.9 % per amplification cycle using standard *Taq* DNA polymerase.[185,186] Moreover, *in vitro* transcription employing T7 RNA polymerase was reported to function with either unnatural base as the templating one.[187]

The *ortho*-substituents relative to the glycosidic bond introduced in the **5SICS**, **MMO2**, and **NaM** unnatural base analogs (*i.e.* sulfur and methoxy groups) were found to be of importance for enzymatic recognition, presumably due to facilitation of H-bonding with the polymerase's active site amino acid side chain residues.[188]

Further insight was achieved by crystallization of elongating deoxynucleic acids in the active sites of a *Taq* DNA polymerase-derived Klenow fragment analog (Klen*Taq*) and *Taq* DNA polymerase (Figure 21).[189] In comparison to a native sequence with a templating d**G** nucleotide and an incoming d**C** TP (Figure 21A/B), striking geometric resemblance of an d**NaM** nucleotide d**5SICS** TP (Figure 21C/D) was observed. This verifies that the enzyme induces natural-like geometry when processing the UBP and might explain its highly efficient replication. In a post-insertion complex showing duplex geometry (Figure 21C/D), however, significant stacking of the d**NaM**:d**5SICS** base pair is prevalent and rather resembles a native mismatch. However, these interactions were considered to beneficially contribute to the overall stability of the UBP.[190]

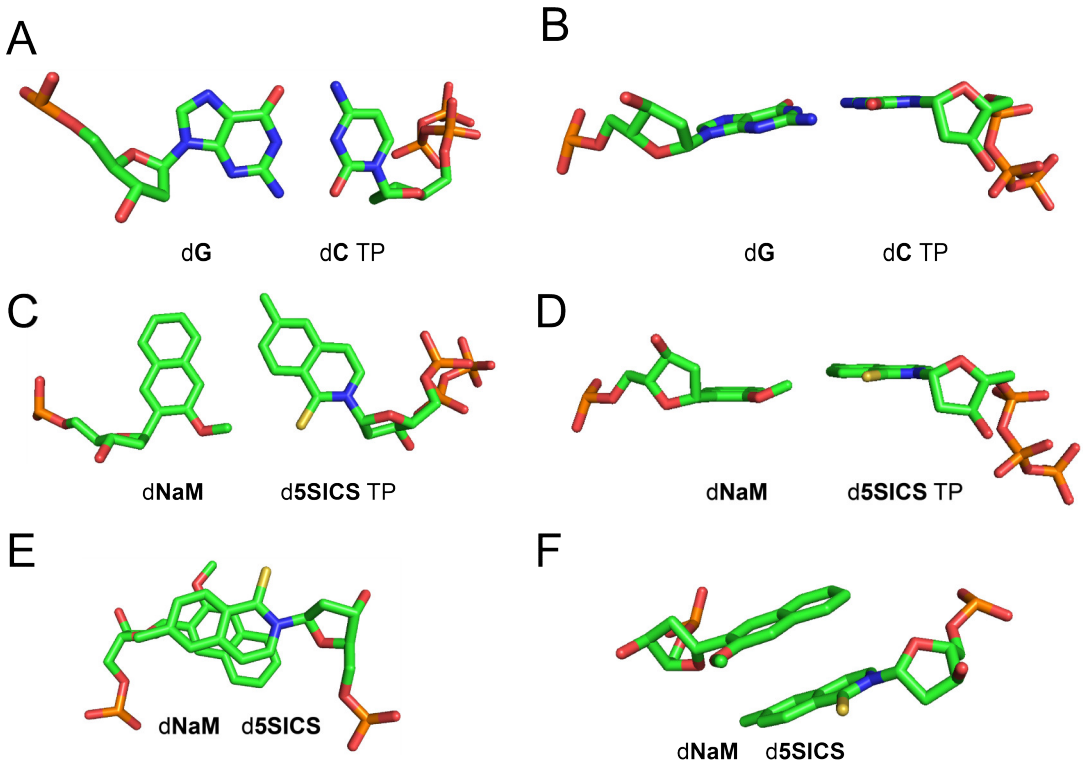


Figure 21. Crystal structures of base pairing modes in top view (left panels) and side view (right panels); A./B. d**G** templating for an incoming d**C** TP in Klen*Taq* DNA polymerase (PDB file: 3rtv)[189], C./D. d**NaM** templating for an incoming d**5SICS** TP in Klen*Taq* DNA polymerase (PDB file: 3sv3)[189]; E./F. d**NaM**:d**5SICS** UBP in a DNA duplex (post-insertion complex with *Taq* DNA polymerase, PDB file: 4c8l)[191].

With an efficiently replicated and transcribed class of novel UBPs approaching the goal of natural-like performance in enzymatic procession, functionalized versions were tested for enzymatic incorporation into DNA and RNA. Notably, fidelity was only slightly decreased and in most cases still ≥99.5 % per PCR amplification cycle. T7 *in vitro* transcription remained largely unaffected by adding linker systems to the unnatural bases **5SICS** and **MMO2**.[192] By this, site-specific introduction of functionalized nucleotides and thus enzymatic or post-enzymatic decoration of nucleic acids with virtually any reporter group is facilitated, which offers profound biochemical and biophysical potential in labeling applications and generation of functionally enhanced molecules such as ribozymes, for example.

Nevertheless, the **5SICS** unnatural base underwent further optimization based on the fact that the intercalating properties of **5SICS**:**NaM** were suspected to hamper extension after enzymatic incorporation. The Romesberg group reasoned that reducing the hydrophobic surface of **5SICS** by removal of its methyl group as well as decreasing the aromatic size would help overcome this issue. Investigation of a panel of analogs identified the **TPT3** nucleobase as most promising candidate for further evaluations (Figure 22A).[193,194]

The hereby established third generation of UBPs by Romesberg and coworkers possessed outstanding PCR amplification efficiency and fidelity in PCR experiments; *e.g.* d**TPT3**:d**NaM** is at least retained with 99.98 % per amplification cycle using *Taq* or One*Taq* DNA

polymerases in an almost sequence-independent manner.[194,195] It was further found, that the **TPT3** scaffold is uniquely tolerant of side modifications: the modified analog **TPT3^{PA}** exhibited an almost identical fidelity of 99.97 % under the same conditions (Figure 22B). Therefore, highly efficient introduction of arbitrary chemical moieties is susceptible for DNA labeling applications and could potentially also be adapted to serve in template-directed transcription reactions.

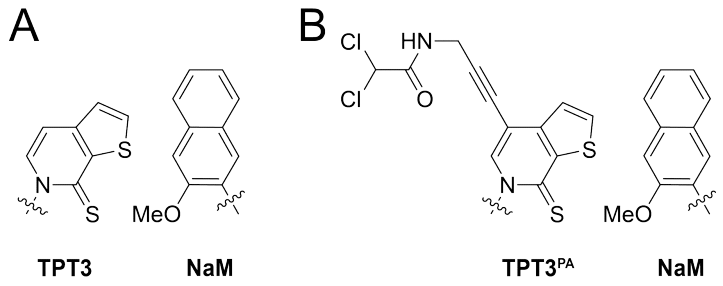


Figure 22. Chemical structures of Romesberg's third-generation UBPs; A. **TPT3**:**NaM**; B. Functionalized version **TPT3^{PA}**:**NaM**; sugar residues are omitted for clarity.

TPT3$^{(PA)}$:**NaM** base pairs do not only exhibit very promising properties, but also their respective nucleic acid building blocks are accessible via convenient and straight-forward chemical syntheses. On the one hand, **TPT3**-derived molecules can be accessed starting from thiophene-3-carboxaldehyde (**18**) (Figure 23A). Knoevenagel condensation with malonate to **19** is followed by a cascade culminating in intramolecular cyclization to give pyridone **20**. After attachment to a suitable ribose scaffold, an optional iodination step creates a branch point for functionalization via standard transition metal chemistry. Further transformations then facilitate generation of the various (deoxy)ribofuranosyl species needed for solid-phase nucleic acid synthesis or enzymatic recognition.

On the other hand, commercially available 2-methoxynaphtalene (**21**) can directly be employed for lithiation and subsequent coupling to sugar moieties, leading to **NaM**-related compounds (Figure 23B).

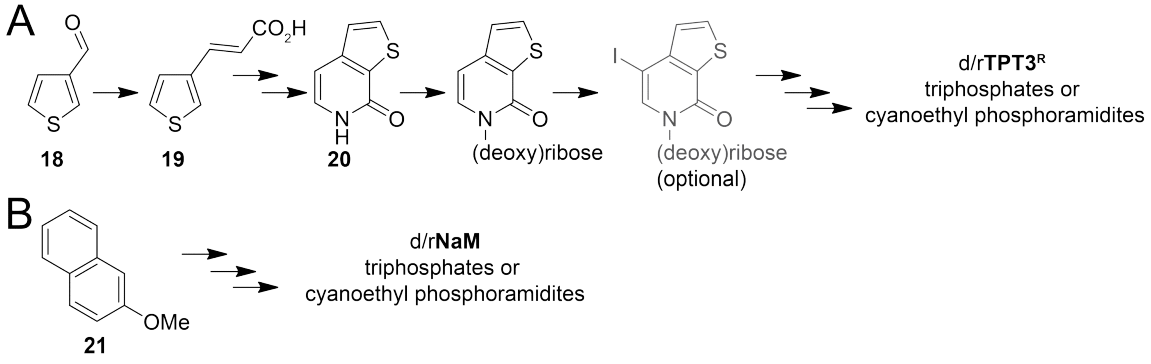


Figure 23. Synthetic schemes towards Romesberg's state-of-the-art UBP building blocks; A. Generation of **TPT3**-derived compounds; B. Access to **NaM**-related species.

Experimentally determined values of the novel UBP **TPT3**:**NaM** could withstand a screening-based approach including 111 unnatural nucleotides (over 6000 base pairing combinations were systematically tested), which verified this combination to be the best available option so far.[193] Notably, several other second- and third-generation UBPs, among them **5SICS**:**NaM** and **5SICS**:**MMO2**, were found to be of similar or only slightly decreased quality; therefore, they are still subject of continuing research in the Romesberg group.

Taken together, these findings strongly suggest an outstanding and general acceptance by enzymes and encouraged for further efforts. For the first time in history, these UBP candidates set a reasonable base to proceed reaching the pronounced long-term goal of developing semi-synthetic organisms (SSOs) that would be able to store and retrieve increased genetic content.

Indeed, pioneering progress was reported employing *E. coli*. The bacteria were transformed with a d**5SICS**:d**NaM**-modified plasmid and provided with unnatural triphosphates. It was shown that they could stably and accurately replicate the augmented DNA while retaining the UBP.[196] Cellular growth was not significantly decreased, showing that these genomic modifications were well-tolerated and non-toxic to the bacteria. Endogenous polymerases are obviously able to accept the respective triphosphate species as substrates, which fascinatingly provides access to orthogonally re-engineered cells.

However, several adjustments on the organism were necessary. For example, the bacteria strain was equipped with an algal nucleoside triphosphate transporter[197,198] to ensure availability of UB TPs as cellular kinase cascades did not sufficiently accept UB nucleosides as substrates due to their high specificity. The replisome of the *E. coli* was further optimized regarding its replication and repair mechanisms increasing the retention of the UBP.[199,200]

Romesberg and co-workers most recently managed to not only store, but also to retrieve increased genetic information in SSOs involving the whole cellular cascade from DNA to protein.[201–203]

Triphosphate transporter-expressing *E. coli* bacteria were transformed with plasmid DNAs containing UBPs. Both a particular mRNA and a tRNA species were encoded in these sequences with an r**NaM**:r**TPT3** base pair placed in the codon-anticodon interaction. The stored information enabled production of a green fluorescent protein (GFP) variant, thus successful translation (and thereby preceding transcription) could be directly tracked by fluorescence readout.[201]

It was demonstrated, that the completely-modified machinery truly facilitated expression of the desired protein species. By contrast, lacking an essential part of the UB-mediated informational flow only led to very low amounts of fluorescence. This was *e.g.* shown by either omitting the UB-modified tRNA or replacing it with near-cognate variants and proved unnatural base-specific use of the central dogma. Evidence of at least 98 % UBP retention

on DNA level and an equally high ratio of protein identity were reported, confirming efficient replication, transcription, and translation involving unnatural codons. Furthermore, the SSO approach was successfully adapted to incorporate non-canonical amino acids into proteins in a similar fashion (Figure 24).[201,202]

While still other UBP combinations are investigated to optimize the properties of six-letter replication, transcription, and translation,[204] Romesberg's fundamental achievements enable an incredibly vast array of fascinating prospective research efforts. Envisioned applications range from site-specific labeling of nucleic acids *in vitro* as well as in living cells to orthogonal transcription networks. Therefore, production of proteins containing unnatural amino acids for augmented SSO systems and therapeutic implementations will be facilitated.

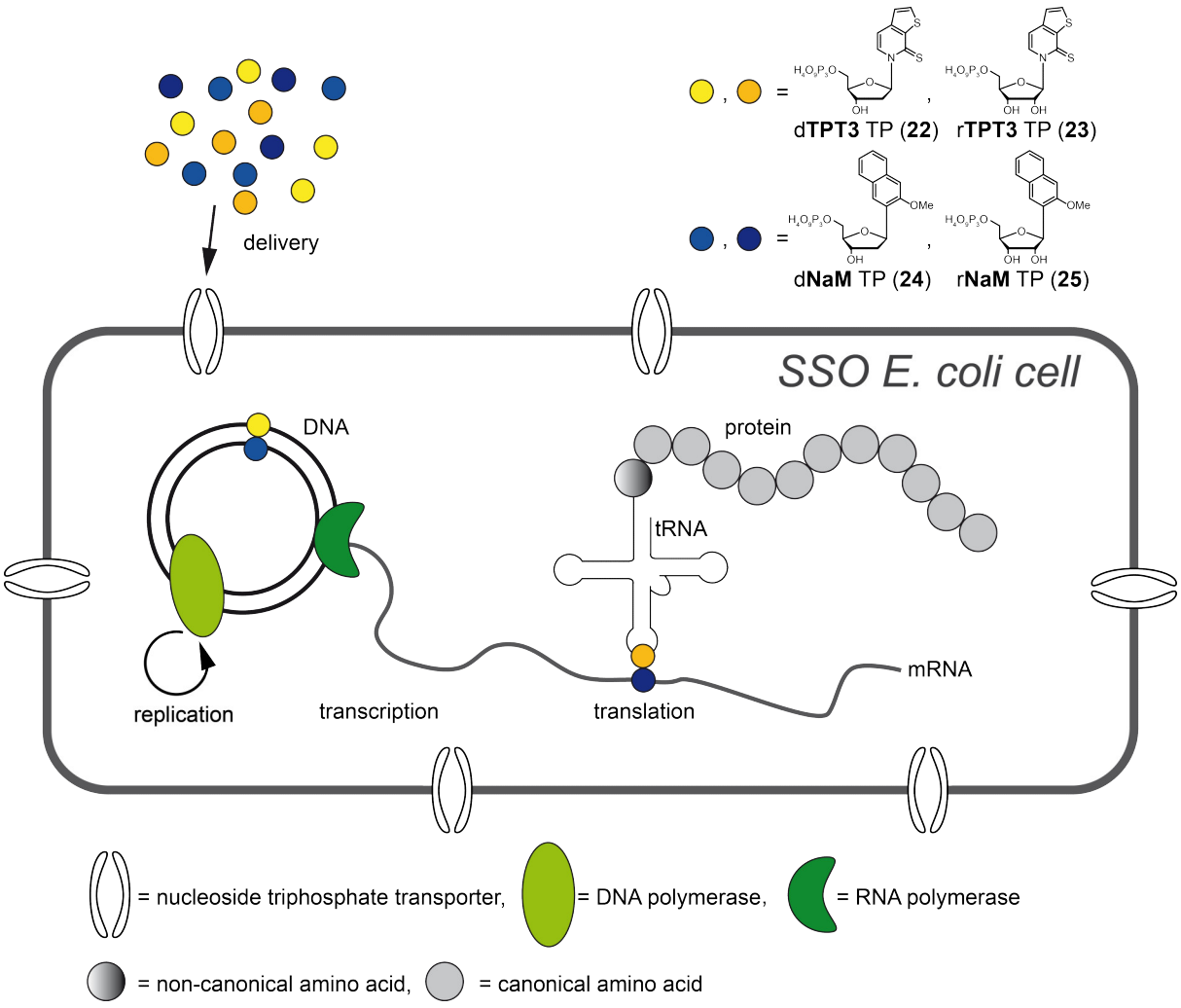


Figure 24. Schematic representation of a fully functional semi-synthetic bacterial organism (SSO) equipped with a transmembrane triphosphate transporter; the SSO is provided with UB TPs d**TPT3** TP (**22**), r**TPT3** TP (**23**), d**NaM** TP (**24**), and r**NaM** TP (**25**) that are delivered into the cell and used by the cellular machinery to replicate d**NaM**:d**TPT3**-modified DNA, transcribe the information including r**NaM**:r**TPT3** nucleotides in codon/anticodon of tRNA and mRNA sequences, and translate the given data into proteins incorporating non-canonical amino acids; reproduced after Dien *et al*.[202]

These inspiring efforts reveal a whole new scenery of accessible nucleic acid modifications. However, little work has been presented regarding the *in vitro* or *in vivo* transcription properties of Romesberg's UBPs.[187,192,205] Strikingly, the most accurately PCR-replicated d**TPT3**:d**NaM** system was not examined for RNA applications by the Romesberg group. Successful *in vivo* experiments suggest pronounced enzymatic recognition on every informational level. The intriguing tolerance of linker systems attached to the **TPT3** scaffold on DNA level[194] makes it an excellent candidate to site-specifically incorporate reporter systems into RNA. Co- or post-transcriptional attachment of reporter groups could avoid impact on sensitive secondary and tertiary structure formation and would give access to data on ncRNA in an unprecedented manner. Also, unlike other techniques currently available, an enzymatic UBP approach would be applicable to RNAs of unrestricted length, providing insights to not-yet addressable targets.

1.5 RNA modifications for biochemical investigations

Intrinsic functionalization of RNA at predefined positions is a prerequisite for a wide range of biomedical and biophysicochemical implementations.[122] Defined tagging can *e.g.* be used for localization within cellular environments, quantification, and isolation of target nucleic acid sequences. Furthermore, site-specific manipulations of desired molecules allow gathering structural insights such as folding phenomena and their dynamics. The ability to control interactions of RNA can facilitate gaining valuable knowledge into cellular pathways, opening doors for biomedical applications, for example by precise regulation of desired features.

1.5.1 Inverse electron-demand Diels-Alder (IEDDA) reactions

Enzymatic approaches are favored to reach the goal of decorating RNA with desired chemical structures. Therefore, a very elegant and convenient methodology is to equip canonical or unnatural nucleotides with small reactive handles. By this, low disturbance of polymerase recognition and structure formation of target sequences is achievable. Following enzymatic procession, further chemical transformations enable coupling with reporter groups. In the past decade, inverse electron-demand Diels-Alder (IEDDA) cycloadditions have raised particular attention upon the scientific world. Most prominently, strained alkenes react with tetrazines in a Diels-Alder retro Diels-Alder cascade (Figure 25). The first reversible cycloaddition step gives a bicyclic intermediate, which upon nitrogen extrusion yields dihydropyridazines that can undergo H-shift isomerization (and eventually oxidation to pyridazines). Several different stereoisomers can result employing asymmetrical substrates. For the sake of simplicity, however, only one species will be depicted throughout this study.

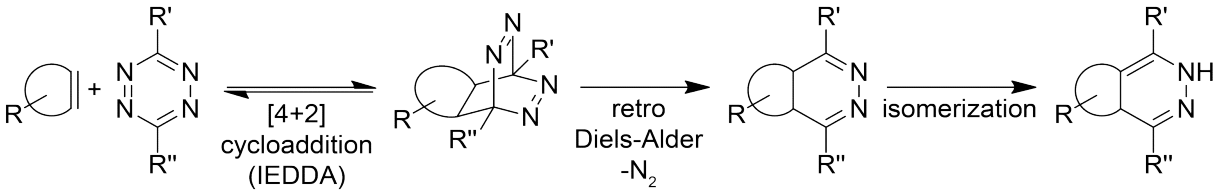


Figure 25. Schematic IEDDA reaction cascade between a strained alkene and a 1,2,4,5-tetrazine.

The application scope of this transformation stretches from material sciences[206–208] to a wide range of biochemistry techniques.[209,210] Offering particularly fast kinetics that are even enhanced in aqueous media[207] in addition to nitrogen as the only evolving and inoffensive byproduct, fulfillment of click chemistry criteria [211] is well-met by IEDDA reactions. Unlike other rapidly converting systems such as active ester chemistry, bio-orthogonality is given since the cycloaddition participants are not influenced by the presence of nucleophiles and the reaction proceeds in reducing aqueous environment as it is given in cells. If desired, reaction rates can be fine-tuned by electronic effects of substituents on the diene and dienophile scaffolds. IEDDA cycloadditions between strained alkenes and tetrazines do not require any additive, making them particularly advantageous for applications in a biological context. Conventional click chemistry involving Huisgen-like cycloaddition of alkynes and azides is performed by copper(I) catalysis, which is known to be cytotoxic[212] and accelerate RNA degradation[213,214].

Significant quenching properties of 1,2,4,5-tetrazines make these reagents especially interesting for chromophoric applications. When conjugated to moieties that emit light at 510–550 nm[215–218], which is the typical magnitude of several commonly used fluorophores such as Oregon Green and boron-dipyrromethene (BODIPY), emission is efficiently suppressed. The tetrazine core is lost upon the IEDDA reaction cascade, which results in turn-on fluorescence (Figure 26A/B). Therefore, biological applications benefit from a low background signal. Turn-on ratios usually are 10–20-fold[217,219,220], however outstanding examples exhibiting several hundred-fold[221] up to 11000-fold[218] increase have been reported. Another advantageous feature is given by the commercial availability of a benzylamine-substituted tetrazine (**26**, Figure 26C), which enables attachment of virtually any chemical residue via active ester chemistry such as *N*-hydroxysuccinimidyl (NHS) species in excellent yields.

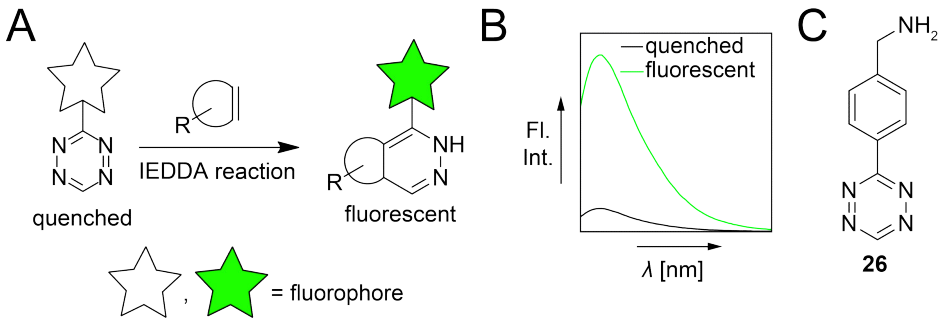


Figure 26. Tetrazine properties for IEDDA reactions; A. Schematic reaction of a quenched tetrazine-fluorophore conjugate with a strained alkene resulting in turn-on fluorescence; B. Representative turn-on fluorescence upon IEDDA reaction; C. Benzylamine tetrazine **26** applicable to functionalization via *e.g.* NHS-ester chemistry.

On the other hand, a set of different strained alkene species is available for IEDDA reactions, most prominently cyclooctenes, norbornenes, and cyclopropenes (Figure 27). Upon the fastest reacting scaffolds are *trans*-cyclooctenes ($k \approx 10^3$ M^{-1} s^{-1})[222–224]. These highly reactive scaffolds have been demonstrated to serve *e.g.* as efficient click partners for site-specific protein glycosylation[225], live cell bioimaging[220], fluorescence labeling of synthetic RNA[226], and even antibody-targeted visualization of lung cancer cells[222]. Unfortunately, the high amount of strain energy makes *trans*-cyclooctenes sensitively prone to isomerization into unreactive *cis* derivatives, which limits their application potential.[227,228] Moreover, commercial availability of *trans*-cyclooctene compounds remains rather poor and expensive. Several preparation methods have been described employing self-constructed liquid flow photoisomerization instruments[223,229,230], however, these still require unusual components that are not accessible in every laboratory.

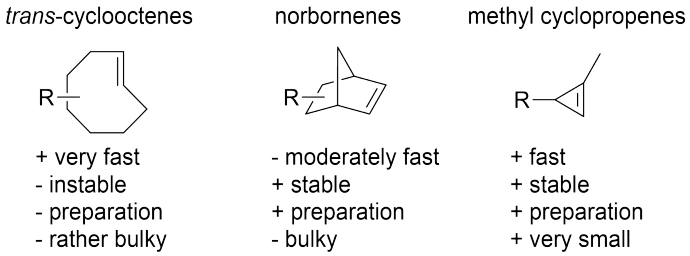


Figure 27. Comparison of strained alkene species for applications in IEDDA reactions.

Norbornene scaffolds are readily available at low costs and have been applied in numerous bio-orthogonal labeling approaches such as site-specific *in vitro* labeling of DNA by means of synthetic incorporation of norbornene-modified nucleotides[231] or enzyme-mediated alkylation of nucleic acids[232], each followed by IEDDA conjugation with functionalized tetrazines. Fluorescent in-cell protein labeling was achieved by the Lemke group using norbornene-modified unnatural amino acids.[233]

Efforts of the Kath-Schorr laboratory proved successful application of a functionalized uridine phosphoramidite (r**U**$^{\mathbf{NOR}}$ CEP, **27**, Figure 28A) in IEDDA cycloadditions on synthetic RNA in mammalian cells (Figure 28B/C) using a tetrazine Oregon Green-conjugate ($^{\mathrm{H}}$**Tet-OG**, **28**), which further supports feasibility of these ultra-mild reactions in a cellular environment.[234]

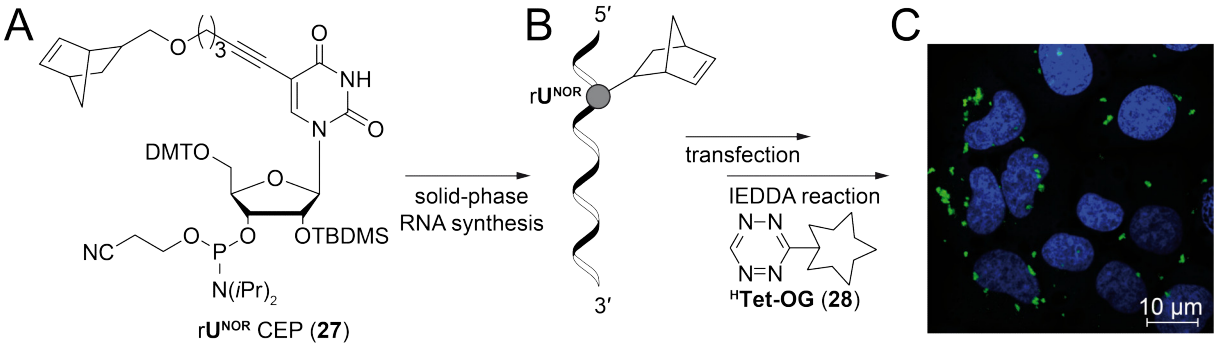


Figure 28. RNA labeling in live mammalian cells via IEDDA cycloaddition[234]; A. Norbornene-modified uridine phosphoramidite r**U**$^{\mathbf{NOR}}$ CEP (**27**), TBDMS = *tert*-butyldimethylsilyl; B. Schematic representation of a chemically synthesized r**U**$^{\mathbf{NOR}}$-modified RNA sequence; C. Confocal fluorescence microscopy image in merge with merge with 4′,6-diamidino-2-phenylindole (DAPI) counterstain visualizing transfected and in-cell fluorophore-labeled RNA by IEDDA click reaction with $^{\mathrm{H}}$**Tet-OG** (**28**).

In addition, S. Kath-Schorr and coworkers demonstrated the first example of a clickable norbornene-modified **TPT3** ribonucleotide (r**TPT3**$^{\mathbf{NOR}}$ TP, **29**, Figure 29A).[235] Employing genetic alphabet expansion transcription based on Romesberg's UBP **TPT3**:**NaM** (Figure 29B) site-specific enzymatic incorporation of r**TPT3**$^{\mathbf{NOR}}$ nucleotides into RNA and post-transcriptional visualization by IEDDA click reaction with $^{\mathrm{H}}$**Tet-OG** (**28**) was shown (Figure 29C/D). This novel approach not only allows for facile and cost-effective labeling of significant RNA amounts, but also spreads the RNA labeling application scope beyond boundaries of chemical solid-phase RNA synthesis and facilitates decoration of potentially arbitrarily long sequences with reporter groups of choice.

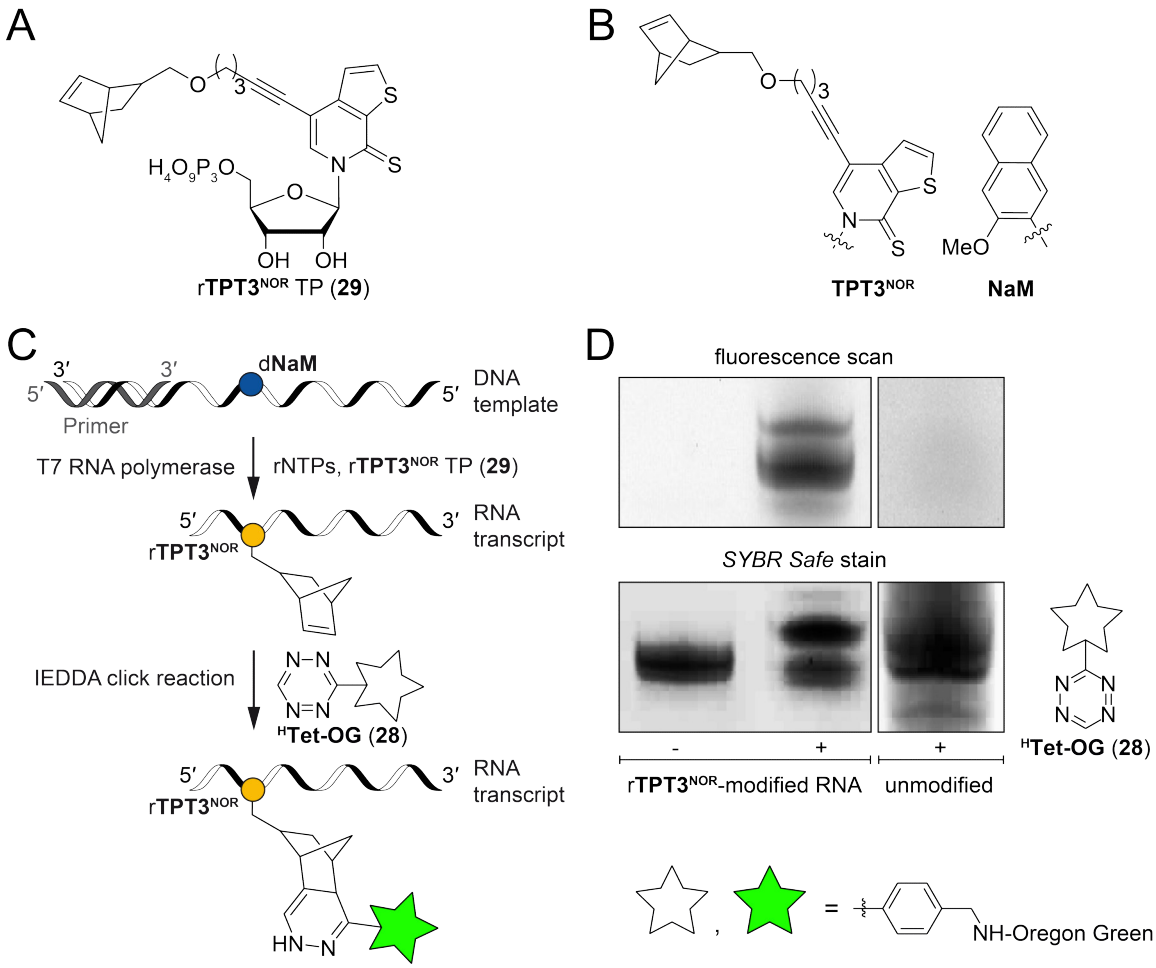


Figure 29. Site-specific labeling of RNA using an expanded genetic alphabet[235]; A. The unnatural triphosphate r**TPT3^{NOR}** TP (**29**); B. Unnatural bases of UBP **TPT3^{NOR}**:**NaM** (sugar residues are omitted for clarity); C. Schematic representation of template-directed RNA labeling via *in vitro* transcription from a d**NaM**-containing DNA in the presence of r**TPT3^{NOR}** TP (**29**) and subsequent IEDDA reaction with tetrazine-fluorophore conjugate **28**; D. Comparative DPAGE analysis of r**TPT3^{NOR}**-mofidied and native RNA transcripts RNA in fluorescence scan and unspecific *SYBR Safe* scan.

However, this UBP system to generate site-specifically labeled RNA might be restricted and remain inapplicable to efficiently generate long RNAs with high yields. Norbornene scaffolds are comparatively sterically demanding[236], which could not only have impact on enzymatic recognition by T7 RNA polymerase during *in vitro* transcription, but also affect secondary structure formation of a target RNA molecules. With focus on ncRNA investigations, correct folding is essential to allow for full functionality and thereby gathering of reliable data. Attachment of a smaller reactive handle to the **TPT3** nucleobase would have several advantages.

Cyclopropene is the smallest IEDDA-susceptive scaffold, methyl-substituted variants are sufficiently stable for biological applications and their high strain energy makes them a better candidate than norbornene moieties that possess rather moderate reaction kinetics (k = 10^{-1}–10^{0} M^{-1} s^{-1} for cyclopropenes vs. k = 10^{-3}–10^{-2} M^{-1} s^{-1} for norbornenes, respectively).[208,237–239] Mini-tags derived from methyl cyclopropene have already been applied in different cell surface probes enabling live-cell imaging by IEDDA reactions with tetrazine-fluorophores[239–241] and protein labeling approaches[242,243].

These fascinating bio-orthogonal applications coupled with their straight-forward synthetic accessibility[244–247] make cyclopropene a very promising and desirable chemical function to further enhance RNA labeling approaches based on the unnatural **TPT3** nucleobase.

1.5.2 Electron paramagnetic resonance (EPR) spectroscopy

EPR spectroscopy is a valuable technique to study structural features and dynamics of biomolecules under solution-state conditions.[248–253] Conceptually similar to nuclear magnetic resonance (NMR) spectroscopy, this method is applicable to molecules containing unpaired electrons. As a prerequisite, the subject of interest thus has to site-specifically carry paramagnetic species. For the analysis of nucleic acids, stabilized nitroxides derived from oxidized piperidine or pyrroline scaffolds are among the most prominent chemical entities employed for this purpose (Figure 30A).

When submitted to a magnetic field, the energy levels of radical spin momenta (m_s = ± ½) split up according to the Zeeman effect (Figure 30B). With a neighboring nucleus of spin I = 1 as given in nitroxides, these levels undergo hyperfine coupling, which results in distinct transition energies ΔE. The correlation $\Delta E = hv = g\mu_B B_0$ (h = Planck constant, v = frequency, g = Landé g-factor, μ_B = Bohr magneton, B_0 = magnetic density) principally allows EPR spectra to be obtained either by constant photon irradiation in a varying magnetic field or vice versa. Usually, frequencies in the microwave magnitude (ca. 1-100 GHz)[254] corresponding to magnetic fields of about 3500 Gauss (G) are applied.

Experiments with continuous (micro)wave irradiation and scanning magnetic force (*cw*-EPR spectroscopy) can give several important information. Generally, existence of particular radical species can be examined by the apparent multiplicity (three peaks for nitroxides). Integration of the obtained signals allows quantifying the spin probe (so-called spin counting). From a qualitative perspective, broadening of peaks and alteration of their relative intensities is associated with immobilization as it would occur when attached to a large biomolecule. Figure 30C exemplarily shows idealized nitroxide *cw*-EPR spectra of different mobility.

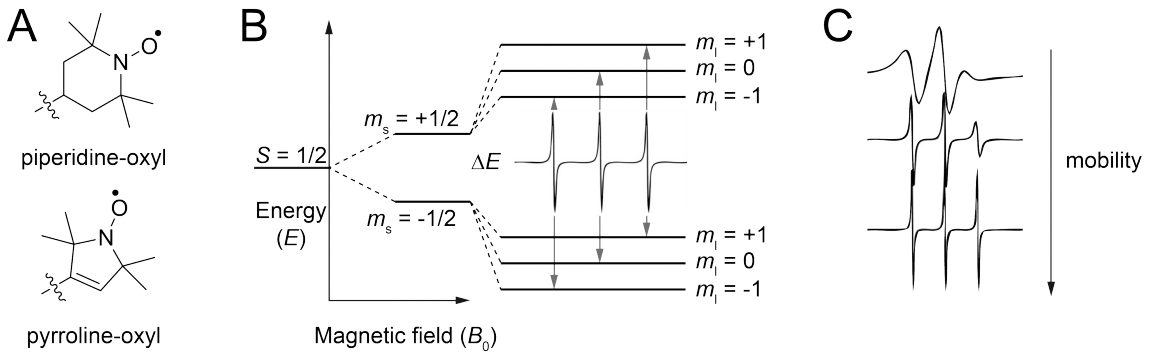


Figure 30. Fundamentals of EPR spectroscopy using nitroxide scaffolds; A. Nitroxide probes derived from piperidines or pyrrolines; B. Energy transitions of nitroxides in a magnetic field; C. Representative *cw*-EPR spectra of differently mobile paramagnetic species; adapted from Jahromy *et al*.[254]

cw-EPR also enables determination of short inter-spin distances (up to ca. 2 nm).[255,256] Moreover, certain sophisticated EPR-based phenomena such as pulsed electron-electron double resonance (PELDOR) can be exploited in spectroscopic experiments to obtain distance distributions between two paramagnetic species even up to ca. 8 nm[257]. For example, dipolar coupling interactions of two nitroxide radicals (Figure 31A) with each other can be addressed. Relaxation times of exited electrons give information about their surroundings, and the amplitude of echo envelopes (Figure 31B) can be used to quantify coupling species. Fourier transformation allows for conversion of raw data into distance vs. probability density plots (Figure 31C). Therefore, EPR spectroscopic techniques can serve as powerful tools to gain precise structural insights in the biologically highly relevant nanometer range.

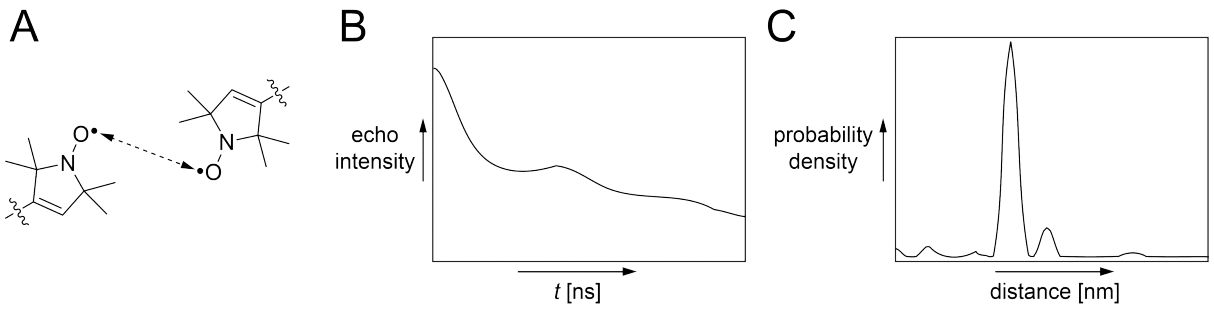


Figure 31. Schematic representation of PELDOR spectroscopic experiments; A. Interactions between two unpaired electrons are measured; B. Time trace of a PELDOR echo envelope; C. Distance distribution after mathematical transformation of raw data; reproduced from Schiemann *et al*[258].

Distance measurements on intramolecular or intermolecular level can involve certain paramagnetic metal ions such as manganese and iron species, which is a particularly interesting feature to investigate biomolecules that naturally exhibit distinct metal binding sites.[259–261] However, if the system of interest does not inherit unpaired electrons, they have to be introduced artificially as probes, termed spin labels. Trityl radicals and nitroxides play

leading roles in biophysical investigations. Studying highly structured RNAs, the sterically less demanding nitroxide species are preferentially used. Sufficient stability towards reduction in aqueous media is provided by α-tetramethyl substitution in pyrimidine and pyrroline oxyl moieties.[253,262,263]

In order to obtain precise data, site-specific incorporation or attachment to the target molecule is essential. Currently, RNA spin labeling techniques mostly rely on solid phase synthesis with nucleotides that bear derivatized nucleobases (Figure 32A).[262,264] Nitroxides were for example post-synthetically attached to functionalized nucleobases using azide-alkyne click chemistry or Sonogashira cross coupling reactions (Figure 32B).[123,127,265] Furthermore, several examples for the introduction of spin labels into nucleic acids via the phosphate backbone (Figure 32C)[266–268] and (deoxy)ribose scaffolds (Figure 32D)[258] have been reported.

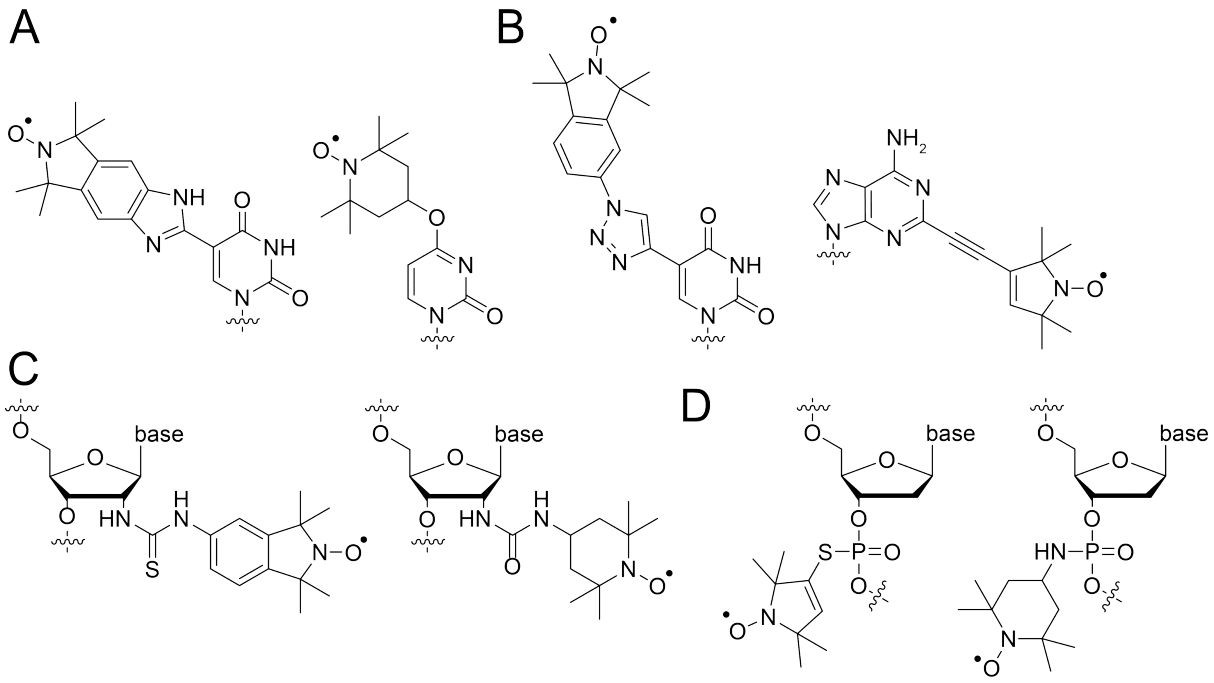


Figure 32. Nitroxide modifications on nucleic acids; A. Functionalized nucleobases[262,264]; B. Post-synthetically attached spin labels to nucleobases[123,127,265]; C. Nitroxides introduced to the 2′-sugar moiety[266–268]; D. Spin labels coupled to the phosphate backbone[258]; residues are omitted for clarity where appropriate.

So far, numerous *in vitro* PELDOR distance measurements on DNA and RNA duplexes[123,127,258,264,269] as well as hairpin motifs[270] and aptamer domains[265,271] demonstrated versatility of EPR spectroscopy to study long-range distances and conformational changes of nucleic acids in solvated state. Even in-cell applications are generally feasible using EPR spectroscopy[259,263,272] and have already been shown to enable distance measurements in nucleic acid model systems in a cellular environment[270,273].

Unfortunately, investigations have been restricted by the limits of chemical solid-phase oligonucleotide synthesis and thus remained inapplicable to larger constructs of interest.
Enzymatic ligation of RNAs has been employed to prepare spin labeled riboswitch sequences of up to 120 nt.[262] Moreover, site-directed introduction of nitroxides into a complexly folded hepatitis V virus genomic RNA section of ≈330 nt could be achieved, however with poor efficiency.[274,275] The approach involved hybridization-based delivery of reactive functions to target sites and subsequent spin label attachment.
Partially overcoming length restriction of site-specific RNA spin labeling for EPR experiments, these techniques utilize laborious multi-step cascades that might result in low yields, potential denaturation of sensitive structural elements, or even decomposition of spin labels.
Expanded genetic alphabet transcription with *e.g.* r**TPT3^{NOR}** TP (**29**), as described in section 1.5.1, combined with post-transcriptional attachment of nitroxide species via IEDDA click chemistry could resolve size limitations of RNA labeling. Direct coupling of nitroxides to the unnatural **TPT3** nucleobase analog would further reduce the effort necessary to introduce spin labels at predefined RNA sites and give facile access to EPR studies on ribonucleic acid molecules of any size and complexity.

1.5.3 Photoswitches

Chemical structures that are able to undergo reversible mechanical and physicochemical transformations upon light irradiation with certain wavelengths are called photoswitches. Controllable interconversion between at least two isomers with distinct properties holds many attractive applications. The possibilities of photoswitch nanoengineering span from optoelectronic devices and stimuli-responsive materials to chemical biology and medicinal chemistry implementations.[276–281]
Many popular devices are based on azobenzene scaffolds (**30**) that are able to reversibly change from the thermodynamically more stable *trans*- into their *cis*-state (Figure 33).

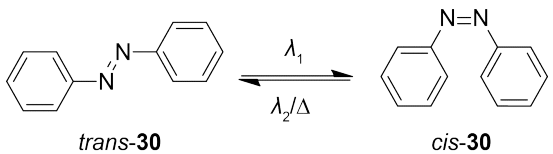


Figure 33. Reversibly *cis*/*trans* photoswitchable diazobenzene (**30**).

Facile synthetic strategies leading to azobenzenes[282] in addition to pronounced structural and polarity differences upon change of conformation make these minimalist entities appealing for versatile approaches. By choice of appropriate substitution patterns

photochemical properties can be tuned shifting the *trans* to *cis* transition wavelength, (λ_1) from the UV range into lower energy spectra.[283]

Spatio-temporal control in biological settings has been achieved to manipulate many essential processes employing photoisomerization of azobenzene motifs. For example, conformations of peptides could be effectively influenced by the introduction of photo-responsive azobenzene moieties linked to amino acids achieving regulation of enzyme activities.[283,284] On nucleic acid level, azobenzenes coupled to sugar subunits were reported *e.g.* reversibly modulating the activity of an RNA-cleaving deoxyribozyme.[285]

Making use of nucleotide surrogates with attached azobenzene-derived residues, the Famulok group presented outstanding efforts on light-controlled DNA hybridization events.[286,287] Their extensive research already enabled the development of a nanocarrier system specifically targeting cancer cells and releasing a therapeutic cargo upon irradiation.[288] Ongoing DNA nanoengineering studies proved light-induced directed molecular movement and paved the way towards non-autonomous molecular machines.[289]

Potential limitations of azobenzene photoswitches are stability issues in reducing environments and unfavored thermal *cis-trans* isomerization.[284]

Diarylethene-based dithienylcyclopentene motifs represent another emerging class of photoisomerizing scaffolds. Generally, the photochromic mode of action is light-controlled reversible ring closure from *open*-**31** into tetracyclic scaffolds related to *closed*-**31** are usually interconverted by UV (λ_1) and visible range absorption (λ_2) (Figure 34).

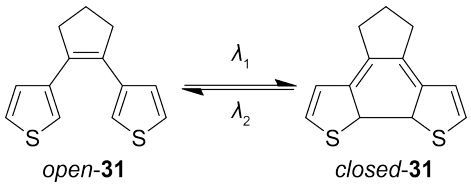


Figure 34. Photo-induced reversible ring-closure of dithienylcyclopentene (**31**).

While being extensively studied with respect to material science contexts[280,290] and optoelectronic applications[291], examples in chemical biology remain rather scarce. Compared to azobenzenes, particularly syntheses of asymmetric photoswitches based on diarylethenes may require more elaborate techniques. However, they can clearly benefit from remarkable reversibility with low fatigue ratio and negligible thermal backswitching.[285]

General applicability of photo-responsive dithienylcyclopentenes under physiological conditions has been demonstrated by several live-cell bioimaging approaches involving fluorescent derivatives.[292,293] Even though diarylethene-based photoswitches arguably exhibit less pronounced three-dimensional change as compared to azobenzenes, rotational degrees of freedom lost during photo-induced interlocking still offer a promising gateway to investigate biological systems in a light-controlled fashion. Successful application of a turn-on

protein kinase inhibitor based on photo-induced diarylethene ring opening by Wilson *et al*.[294] strongly supports sufficient steric differences relevant for enzymatic discrimination.

Furthermore, the Jäschke group invented a set of photoswitchable nucleosides based on natural nucleobase scaffolds. For example, deoxyadenosine-(d**A**)-derived compound **32**[295] (Figure 35A) could reversibly undergo efficient electrocyclic rearrangements using short UV/visible light (VIS) pulses without recognizable loss of activity. The closed form exhibited increased absorption properties in the fluorescence spectral range, making the compound attractive to serve as conditional quencher. Furthermore, deoxyuridine-(d**U**)-based photoswitchable nucleoside **33** (Figure 35B) was converted into its CEP and incorporated into solid phase-synthesized short DNA antisense primers containing the T7 RNA polymerase promoter sequence.[296] Reversible light-controlled regulation of T7 *in vitro* transcription efficiency was impressively shown, suggesting powerful applications for manipulation of nucleic acid-protein interactions.

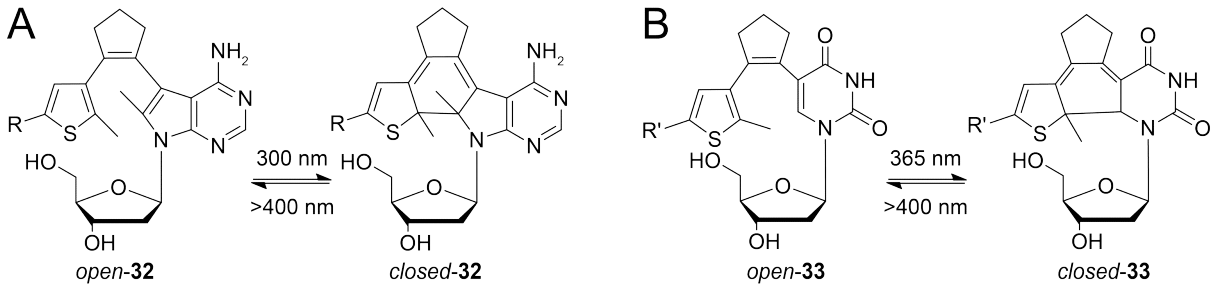


Figure 35. Photoswitchable deoxynucleosides by the Jäschke group[295,296]; A. Nucleoside **32** based on d**A**, R = pyridinyl; B. Nucleoside **33** based on d**U**, R' = phenyl.

These findings deeply inspire for prospective research in light-induced nucleic acid modifications. Attachment of diarylethene-based photoswitches to artificial nucleobase analogs such as the **TPT3** thienopyridone building block belonging to Romesbergs's UBP **TPT3**:**NaM** could offer opportunities to site-specifically incorporate such functions into DNA and RNA. Manipulation of RNA interactions could be possible beyond the limitations of solid-phase synthesis. This pioneering work would set the base to study ncRNAs in a spatially and temporally controlled way and set the base to evolve novel ribozymes with switchable activities.

2 Research objectives

The biological roles of functional non-protein coding RNAs cannot be overestimated. However, knowledge about their abundance, architecture, and impact dimension is still significantly underdeveloped. Functionality of these biomacromolecules is determined by sensitive structure formation. Non-disturbing investigation of such molecules remains a challenging task.

Overcoming restrictions of currently available RNA labeling techniques, this study aimed to establish new approaches allowing for facile and site-specific introduction of reporter groups into highly structured RNA molecules with minimal impact on their native state.

Implementation of an expanded genetic alphabet, particularly involving Romesberg's UBP **TPT3**:**NaM**, was envisaged to facilitate preparation of long modified RNA constructs via template-directed T7 *in vitro* transcription reactions.

Primary focus was set on the improvement of former efforts on a norbornene-functionalized **TPT3** nucleotide for *in vitro* fluorescence labeling of short RNA molecules. Replacement of the rather bulky alkene moiety attached to the unnatural nucleobase scaffold by a smaller and more reactive cyclopropene handle was aspired. Synthesis of a novel cyclopropene-functionalized **TPT3**-based ribonucleoside triphosphate building block promised excellent enzymatic recognition and procession as well as minimal steric impact on target RNA transcripts.

The cyclopropene linker was intended to serve in post-transcriptional IEDDA click reactions with tetrazine derivatives, giving straight-forward access to target RNAs site-specifically conjugated to arbitrary reporter groups such as fluorophores under bio-orthogonal conditions. Characteristics of enzymatic incorporation efficiency and specificity were necessary to be addressed in a detailed manner using *e.g.* mass spectrometry, gel electrophoresis, and fluorescence-based assays.

Consecutively, establishment and application of methods to prepare site-specifically modified RNA sequences of any length by introduction of novel unnatural **TPT3** nucleotides was conceived. Modification of model functional RNA systems such as ribozymes with unnatural nucleotides was planned to elucidate and exclude disturbance of secondary structure formation. Syntheses of the **TPT3**:**NaM** UBP deoxynucleoside triphosphates were performed for this purpose.

Furthermore, achievement of site-specific RNA spin labeling using expanded genetic alphabet technology was aspired in a similar fashion. Ultimately, development of a nanometer-range ruler for distance measurements in complexly folded RNA molecules under solution conditions was intended.

Elegant access to precise information about structural dynamics of RNA beyond the limits of chemical solid-phase synthesis would thus be provided. Insights into general feasibility and resolution of the proposed method using EPR spectroscopic methods were expected by analysis of short spin labeled self-complementary duplexes. Subsequent application of UBP-mediated spin labeling to a larger RNA molecule with distinct three-dimensional structure was envisioned to confirm versatility of the novel technique.

Based on the **TPT3** scaffold, studies towards a diarylethene-based photoswitchable unnatural nucleotide were envisaged. Successful identification of a reliable multi-step synthetic route to prepare appropriate RNA building blocks was proposed to prospectively enable either enzymatic or synthetic incorporation. Site-specific modifications of RNA sequences with a chemical entity that reversibly undergoes conformational change upon light irradiation are assumed to facilitate spatial and temporal control of RNA interactions, thus creating unprecedented opportunities to study structural dynamics of complex molecules.

Application of the UBP **TPT3**:**NaM** for RNA investigations represents pioneering work. Thus, general aspects such as site-specificity of enzymatic incorporation and possibilities to reversely transcribe modified sequences were to be addressed. Chemical access to further **TPT3**-related ribose building blocks of unnatural base scaffolds as well as data on promises, challenges, and limitations of an expanded genetic alphabet were required throughout this study.

This work intended to make a major contribution extending the scope of site-specific enzymatic RNA labeling techniques by studying and applying extended genetic code approaches. Investigation of highly structured transcript molecules was proposed via intrinsic and non-interfering introduction of reporter modifications using a class of sophisticated unnatural base pairs.

3 Results and discussion

Various achievements based on an extended genetic code were accomplished during the course of this study. Synthesis of a novel cyclopropene-functionalized ribose triphosphate as counterpart for an unnatural base pair is presented in section 3.1.1. Section 3.1.2 includes development and thorough investigations to apply this compound in several powerful site-specific RNA labeling techniques via *in vitro* transcription with an expanded genetic alphabet. Nanometer-range distance measurements on RNA transcripts have been achieved using EPR spectroscopy. Efforts involving post-transcriptional or direct introduction of spin labels during *in vitro* transcription with an unnatural base pair are described in section 3.2.1 and 3.2.2, respecively.

Furthermore, the synthetic route towards a reversibly photoswitchable ribonucleoside based on an unnatural base scaffold was elaborated and conducted to a large extent (section 3.3).

Section 3.4.1 contains syntheses of several DNA and RNA building blocks for further enzymatic applications with an unnatural base pair.

Insights into enzymatic recognition of unnatural bases via enzymatic digestion of RNA transcripts as well as a reverse transcription approach are shown in sections 3.4.2 and 3.4.3.

The overall progress of this work will likely enhance future efforts in the field of non-coding RNA research and contribute to scientific work involving an augmented genetic alphabet.

3.1 Site-specific RNA labeling using the novel cyclopropene unnatural nucleotide r**TPT3^{CP}** TP (**34**)

Enhancing the potential and applicability of formerly developed norbornene-modified r**TPT3^{NOR}** TP (**29**)[235], synthesis of a novel **TPT3**-derived ribofuranosyl triphosphate, r**TPT3^{CP}** TP (**34**), was envisaged. Bearing a small reactive methyl cyclopropene handle, outstanding incorporation efficiency during *in vitro* transcription from d**NaM**-modified DNA templates and fast kinetics in IEDDA cycloadditions with tetrazine derivatives were expected. This would greatly expand site-specific RNA labeling methods to attach virtually any reporter group at predefined positions under particularly mild and bio-compatible conditions with potential in-cell applicability.[209,237–240,242,297]

Based on recent research in the Kath-Schorr group[235,298], the synthetic route towards r**TPT3^{CP}** TP (**34**) was completed and optimized. Incorporation dynamics and efficiency were addressed employing short exemplary transcripts. Furthermore, approaches to facilitate r**TPT3^{CP}** labeling of RNA sequences independent of their length were sequentially established and evaluated. The following sections will address all of these topics in sound detail.

3.1.1 Synthesis of r**TPT3^{CP}** TP (**34**)

The synthesis of cyclopropene-derivatized r**TPT3^{CP}** triphosphate **34** took place in analogy to the assembly of r**TPT3^{NOR}** TP (**29**) reported earlier[235], which was based on the synthetic route to a modified **TPT3** nucleobase analog (**TPT3^{PA}**, see Figure 22, section 1.4.3) as published by Li *et al*.[194]

Briefly, the corresponding nucleoside **35** was retro synthetically fragmented into nucleobase analog precursor **20,** a cyclopropene alkyne linker (**36**), and protected ribose **37** (Figure 36). Preparation and coupling of these building blocks was planned in a straight-forward manner partly adapted from literature[194,239].

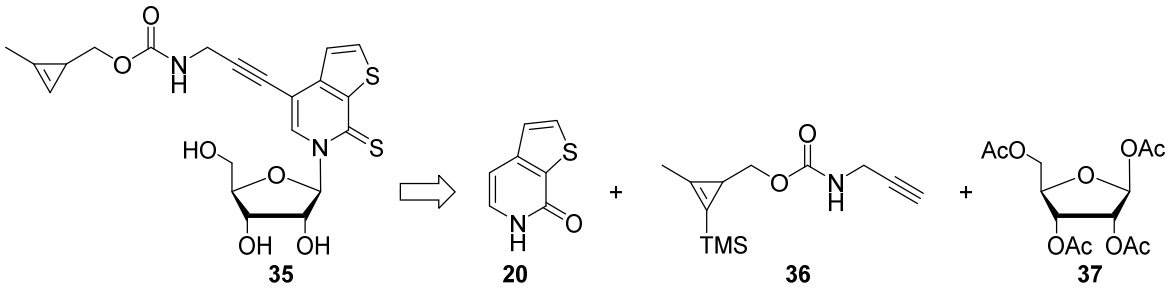


Figure 36. Retrosynthetic considerations providing access to r**TPT3^{CP}** nucleoside **35**.

3.1.1.1 Synthesis of cyclopropene linker 36

Access towards the novel methyl cyclopropene propargylamine linker **36** was envisaged in three steps adapting synthetic procedures from Petiniot *et al.*[299] and Yang *et al*.[239] (Figure 37). Fusion of trimethylsilyl (TMS) propyne (**38**) and ethyl diazoacetate (EDA, **39**) was supposed to give TMS-protected methyl cyclopropene ethyl ester **40**. Subsequent reduction to the corresponding primary alcohol **41** was to be followed by carbamoylation employing propargylamine (**42**) to provide title compound **36**.

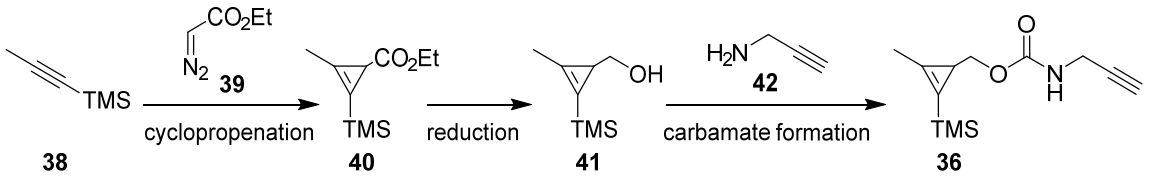


Figure 37. Schematic synthesis of cyclopropene carbamate **36**.

Synthesis of **36** was carried out as suggested above. In a first step the methyl cyclopropene scaffold **40** was prepared in 90 % yield by rhodium(II)-catalyzed addition of ethyl diazoacetate (**39**) to trimethylsilyl propyne (**38**) (Figure 38). Reduction of the ethyl ester function to the corresponding primary alcohol with diisobutylaluminium hydride (DIBALH) led to 93 % **41**. A two-step reaction involving carbonyldiimidazole (CDI) and propargylamine (**42**)

then afforded 85 % of the cyclopropene carbamate alkyne linker **36**. An overall yield of 71 % was achieved over four consecutive steps, providing easy and straight-forward production of a versatile cyclopropene alkyne linker that is attachable to a vast variety of suitably halogenated substrates under *e.g.* Sonogashira coupling conditions.

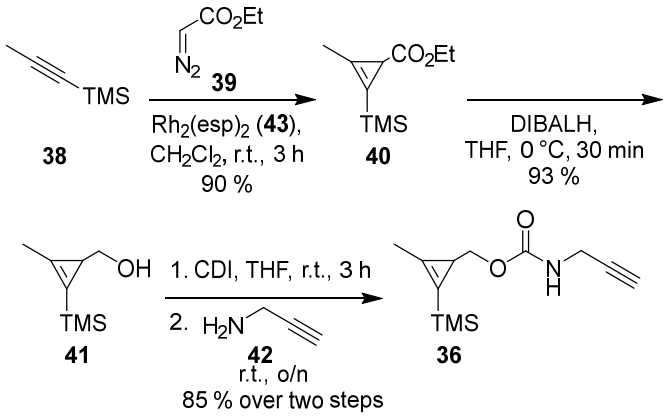


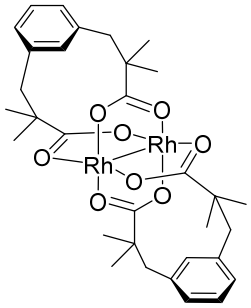
Figure 38. Synthesis of cyclopropene linker **36**.

Cyclopropenation[246,299]

The cyclopropenation step was tested with two different rhodium-based catalysts, rhodium acetate dimer ($Rh_2(OAc)_4$) and bis[rhodium(α,α,α′,α′-tetramethyl-1,3-benzenedipropionic acid)] ($Rh_2(esp)_2$, **43**, Figure 39), both of them leading to comparable reaction success.

Though the rhodium acetate dimer is reported to be a good cyclopropenation catalyst[244,245,299,300], formation of an unknown byproduct was found which could not easily be separated by column chromatography or distillation[298]. The following steps (*i.e.* reduction and carbamoylation), however, allowed separation from the unwanted compound. Therefore, synthetic success can be achieved with this route, neglecting impurity of cyclopropenyl ethyl ester **40**. $Rh_2(esp)_2$ (**43**) in contrast was found to be a far more reactive rhodium species.[246,301] Cyclopropenation proceeds smoothly while extensive nitrogen extrusion can be observed. No byproducts were formed and the addition of ethyl diazoacetate (**39**) was conducted in less time than with $Rh_2(OAc)_4$ as a catalyst. **43** in turn has considerable drawbacks on its own. Commercial availability is not as prevalent as the rhodium(II) acetate dimer and **43** is of considerably higher pricing ($Rh_2(esp)_2$, **43**: 492.00 € g^{-1}, *Sigma Aldrich*, accessed 2019/03, $Rh_2(OAc)_4$: 244.00 € g^{-1}, *Alfa Aesar*, accessed 2019/07).

Notably, product purification involves filtration of the crude reaction over Celite®, holding back the catalyst. This, however, was not possible in every case. A certain batch dependency was observed which is supported by differing appearance of the catalysts, ranging from light-green to intense turquoise/blue-colored powders.



43

Figure 39. Chemical structure of highly reactive rhodium-based catalyst $Rh_2(esp)_2$, **43**.

Generally, both catalysts employed in this study are conceivable options for the cyclopropenation step in the cascade to assemble cyclopropene linker **36**. Aiming on high purity of cyclopropene ester **40**, the use of catalyst **43** is advised. With focus on budget, the rhodium acetate dimer facilitates sufficient reaction success for downstream applications if handled properly.

The cyclopropene ester **40** appears to be volatile, thus special care has to be taken concentrating crude reaction mixtures as well as chromatographically obtained fractions containing the product (100 mbar, 45 °C). It seems, however, that a certain share is still co-evaporated during these processes, preventing quantitative isolation of **40** despite full conversion of starting material **39**.

Ester reduction[239]

Converting the ethyl ester function of **41** to the primary alcohol proceeded smoothly with a commercially available solution of DIBALH in tetrahydrofuran (THF). As found earlier[298], this complex metal hydride is the reagent of choice since more reactive hydride donors like $LiAlH_4$ tend to react with the strained cyclopropene ring and give cyclopropane derivatives as byproducts[244] or major components[302]. DIBALH on the other hand is primarily known to convert esters into aldehydes using apolar solvents. When THF is employed as the solvent, it is able to coordinatively stabilize the transition state in a second hydride donation process[303] thus facilitating mild synthesis of cyclopropenyl alcohol **41**. During workup the slight excess DIBALH was quenched by addition of a saturated potassium sodium tartrate (Rochelle salt) solution which is most commonly used in literature procedures.[304] A short time after the addition, however, an aqueous gel formed, which had to be mechanically crushed in order to ensure proper extraction. This is not a particularly desirable feature for a straight-forward workup and prospective attempts could involve other mild quenching reagents.

Cyclopropenyl alcohol **41**, similar to its precursor **40**, is easily lost during *in vacuo* concentration and careful evaporation (100 mbar, 45 °C) is essential in order to obtain maximum yields.

Carbamate formation[239,305]

The two-step conversion with CDI and subsequent reaction with propargylamine (**42**) readily gave good yields of **36**. In former studies[298] the activated carbonic acid of **41** (*i.e.* the acylimidazolyl mixed anhydride) was reacted with **42**, affording ~70 % of product. The yield was improved to 85 % with an increased amount of CDI (1.1 equivalents, eq.). Literature even states near-quantitative isolation adding Hünig's base (diisopropylethyl amine, DIPEA).[305]

3.1.1.2 Synthesis of thienopyridone 20

Synthesis of the **TPT3** nucleobase analog precursor **20** (Figure 40) was planned involving procedures by Cipollina and coworkers[306] as well as a patented description by Ewing *et al*.[307]

Starting from thiophene-3-carboxaldehyde (**18**), standard Knoevenagel condensation with malonic acid (**44**) was scheduled enabling easy access to (2*E*)-3-(3-thienyl)-2-propenoic acid (**19**). Subsequent conversion into its corresponding acid azide via intermediate *in situ* activation of the carboxylic acid function with *e.g.* ethyl chloroformate (**45**) was intended. Curtius rearrangement into its isocyanate was supposed to facilitate intramolecular ring closure[306,308,309] giving pyridone **20**.

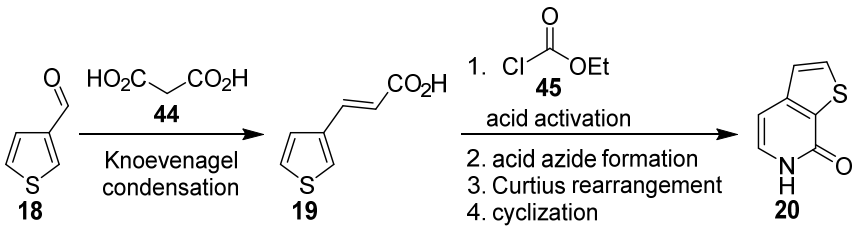


Figure 40. Schematic synthesis of thienopyridone **20**.

As intended, commercially available thiophene-3-carboxaldehyde (**18**) could be converted into acrylic acid **19** by Knoevenagel condensation with malonic acid (**44**) via reflux in pyridine (py) and catalytic amounts of piperidine with a near-quantitative yield of 95 % (Figure 41). A three-step one-pot reaction then afforded 75 % of target compound **20**. For this, *in situ* generation of a mixed anhydride **46** with ethyl chloroformate (**45**) under basic conditions was followed by intermediate transformation into acid azide derivative **47** using NaN_3. **47** was thermally decomposed at elevated temperatures into the corresponding isocyanate which eventually intramolecularly cyclized giving the title compound. Starting from **18**, thienopyridone **20** was prepared with a 71 % overall yield. Taking into consideration that commercial sources offer the product starting from ~600 € g^{-1} (*Sigma Aldrich*, accessed

2019/03), the optimized protocol presented in this study represents a proper route to multi-gram synthesis of **20** employing low-cost starting materials and reagents.

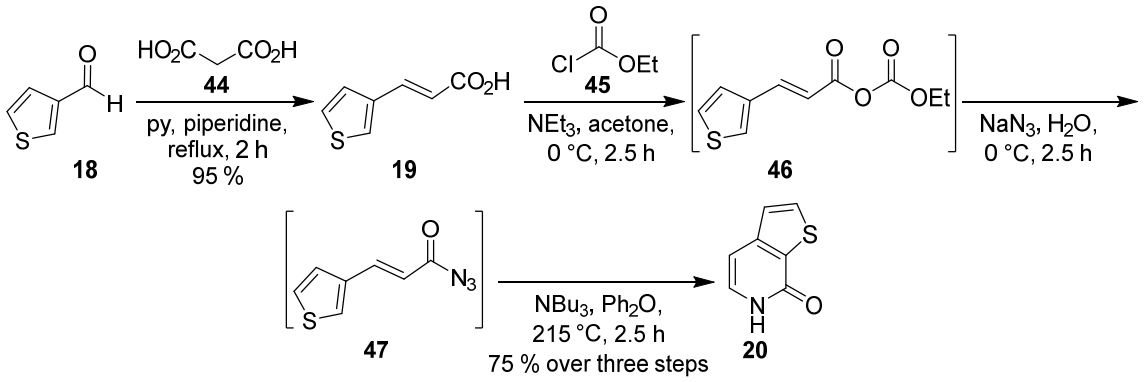


Figure 41. Synthesis of pyridone **20**.

Knoevenagel condensation[194,306,310–312]

Formerly, the Knoevenagel condensation product **19** was precipitated and collected by filtration. As precipitations tend to be non-quantitative in many cases, the protocol was changed and the crude mixture was extracted with ethyl acetate (EtOAc) despite the occurring solid. No significant formation of byproducts is observed during this reaction, extraction thus facilitates maximum isolation of the target substance.

Pyridone preparation[194,307,308,312]

During generation of **20**, the acid azide derivative **47** was intermediately isolated by precipitation from water. Here, proper desiccation is essential, as the powdery substance is subsequently added to tributylamine (NBu_3) in phenyl ether (Ph_2O) at 215 °C. Residual water hampers reaction success and causes safety risks due to deflagration effects. Overnight freeze-drying overcomes this issue and is strongly recommended.

3.1.1.3 Assembly of cyclopropene-modified triphosphate 34

Synthesis of a novel cyclopropene carbamate-modified ribonucleoside triphosphate bearing the unnatural **TPT3** nucleobase analog was suggested adapting procedures from Vorbrüggen and Bennua[313], Li *et al.*[194], Antonini *et al.*[314], Yang *et al.*[297], as well as Srivatsan and Tor[130].

Generally, production of nucleoside precursor **48** was intended via attachment of thienopyridone derivative **20** (see section 3.1.1.2) to acyl protected ribose unit **37** applying the Vorbrüggen reaction (Figure 42). Iodination of the nucleobase analog yielding **49** was envisaged to afford susceptibility towards Pd(0) cross-coupling derivatization reactions. Development of ribosyl r**TPT3** nucleoside precursor **50** by thionation of the pyridone carbonyl

oxygen was supposed to be followed by removal of ribose acyl groups under basic conditions. Coupling of TMS-protected cyclopropene alkyne linker **36** to iodinated r**TPT3** nucleoside **51** was planned under Sonogashira reaction conditions. Fluoride-mediated TMS-deprotection of the resulting functionalized nucleoside **52** was proposed to yield the r**TPT3^{CP}** nucleoside (**35**), which gives access to the desired cyclopropene-modified r**TPT3^{CP}** TP (**34**) by application of Ludwig's triphosphate synthesis strategy[315].

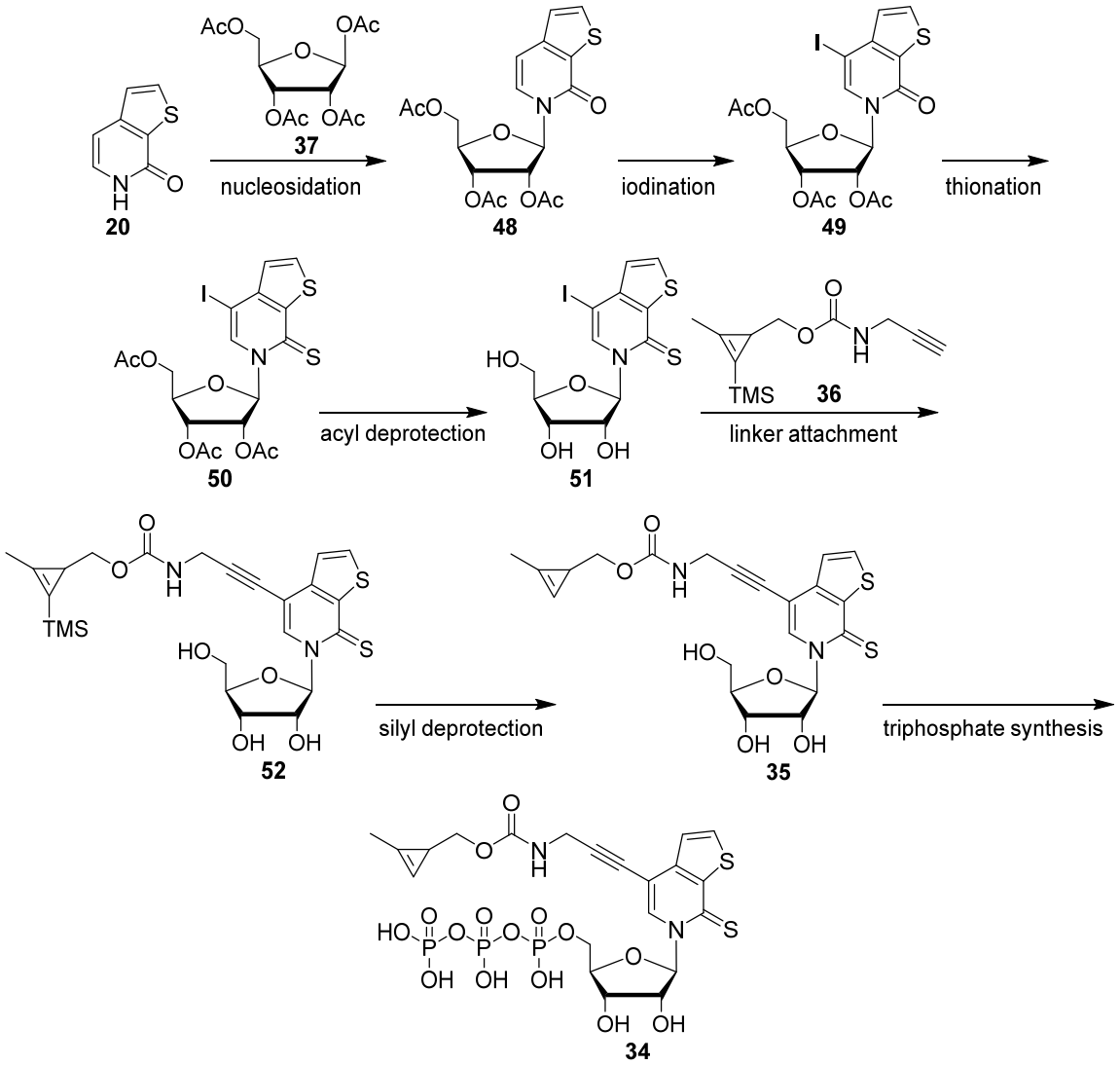


Figure 42. Schematic synthesis towards cyclopropene-modified ribonucleoside triphosphate r**TPT3^{CP}** TP (**34**).

This synthetic route allows for the production of a novel unnatural base triphosphate with inherent IEDDA reactivity in a well-tested and reliable manner. After a brief overview of reagents and product yields, individual steps will be discussed in detail.

Nucleobase analog **20** was first coupled to peracylated D-ribose **37** under Vorbrüggen (or advanced silyl Hilbert-Johnson) conditions[313], obtaining **48** in 78 % yield (Figure 43). An iodination step with iodine monochloride afforded 43 % of **49**, which was subsequently thionated using Lawesson's reagent (**53**) with a yield of 88 %, receiving compound **50**.

Removal of acyl protecting groups from the sugar residue employing a methanolic ammonia solution quantitatively led to nucleoside **51**. Cyclopropene linker **36** was attached under Sonogashira conditions to give **52** in good yield (77 %). After TMS deprotection of the cyclopropene residue with tetrabutylammonium fluoride (TBAF), 77 % of purified cyclopropene-modified nucleoside **35** were isolated. Final three-step conversion into the corresponding triphosphate with $POCl_3$, subsequent reaction with tributylammonium pyrophosphate, $(Bu_3NH)_2PPi$, and buffered hydrolysis, gave 31 % of r**TPT3^{CP}** TP (**34**). The presented nine-step synthetic route facilitated generation of **34** starting from pyridone derivative **20** in ca. 5 % overall yield.

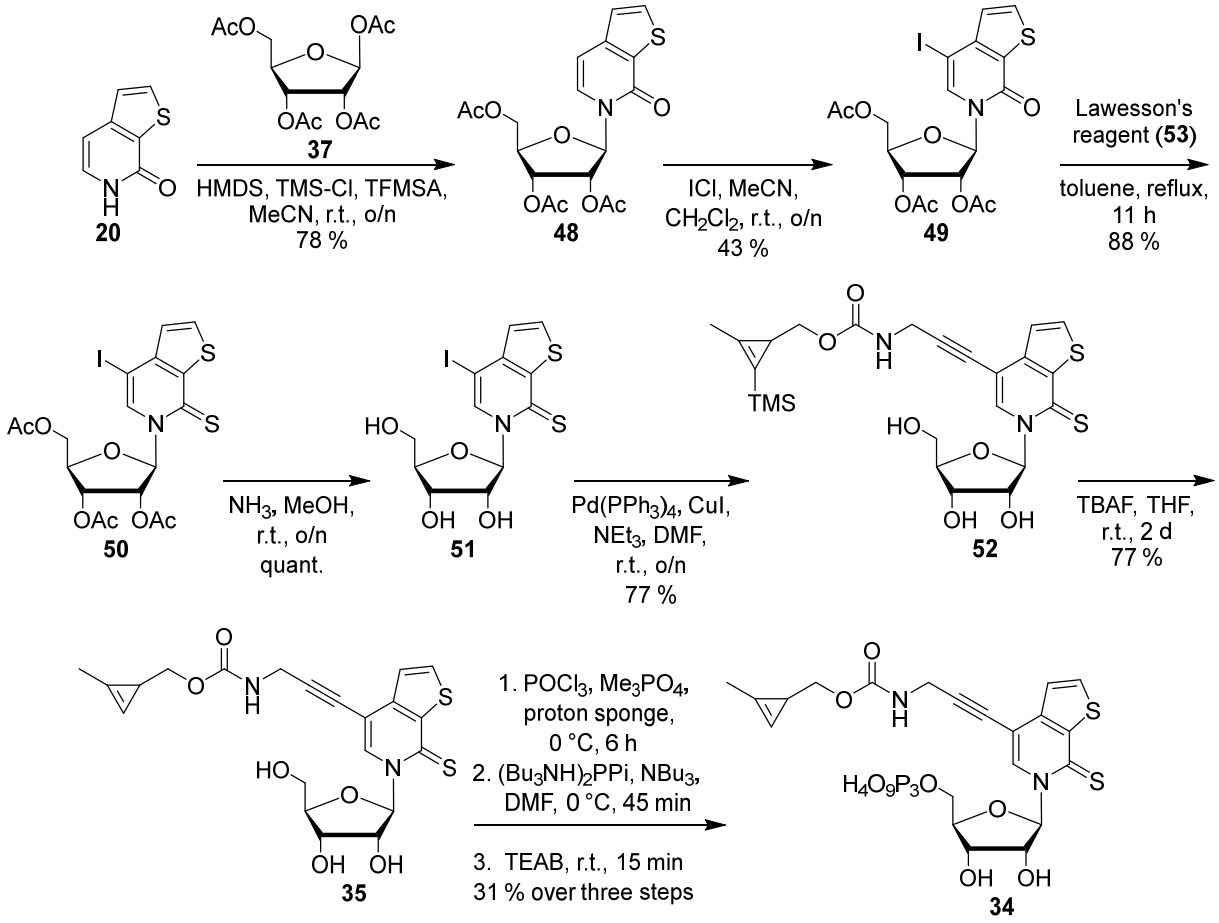


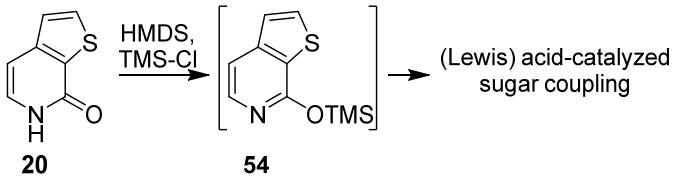
Figure 43. Syntesis of r**TPT3^{CP}** TP (**34**).

Nucleoside formation[313,316,317]

The Vorbrüggen or silyl Hilbert-Johnson reaction to give 2′,3′,5′-acylated ribose adduct **48** was improved from formerly 56 %[235] to 78 %. One major change in the reaction protocol was a prolonged reaction time with hexamethyldisilazane (HMDS) and trimethylsilyl chloride (TMS-Cl) to intermediately give the silylated species **54** (Figure 44), readily reacting with the protected ribose. Furthermore, the temperature of trifluoromethanesulfonic acid (TFMSA)

addition was changed from 0 °C to room temperature, which is assumed to be an additional improvement for the sugar coupling.

Despite the presented yield, addition of **20** to **37** is still improvable as reactions rather result in heterogeneous success with an average of ~50 % and a lower limit of ~30 % isolated product. However, in many cases unreacted pyridone derivative **20** precipitated during workup and could be isolated to enable a total recovery of up to 99 % (sum of product yield and recollected **20**). This suggests a certain lack of reactive pyridone species and further optimization would be desirable. Probably, equivalents of HMDS and/or TMS-Cl could be increased to give full conversion to silylated species. Additionaly, intermediate isolation of **54** could help improving coupling success, as excess HMDS could negatively impact the catalyst's performance[318]. Further significant factors on the reaction success include application of ultrapure and thoroughly pre-dried ribose compound **37**[317] and the choice of solvent (nitromethane and 1,2-dichloroethane could serve as potential alternatives to MeCN)[313]. Also, a series of different Friedel-Crafts catalysts such as trimethylsilyl perchlorates and triflates are known to be effective[317] and could be tested in this particularly complex step, possibly further improving the experimental outcome.



20 **54**

Figure 44. Schematic *in situ* formation of silylated species **54** from pyridone **20** with HMDS and TMS-Cl.

Iodination[194]

Iodination of ribose adduct **48** with iodine monochloride is of major concern regarding the overall yield in the multi-step synthesis of r**TPT3^{CP}** TP (**34**). Former work on this project revealed that reaction success with ICl crucially depends on the solvent choice.[298] In apolar solvents such as dichloromethane significant chlorination[319,320] (approx. one quarter) was observed and hampered a satisfying reaction yield (up to 38 % overall halogenated products). More polar acetonitrile (MeCN) as solvent overcomes byproduct formation, however, still gives moderate yields of ~40 % **49**.

It is expected, that the mechanism of this step involves intermediate formation of radical iodine atoms[320] (Figure 45) that could recombine with each other leading to molecular iodine, which could be insufficiently reactive to iodinate **48**. Increasing the equivalents of ICl from 1.1 to 1.5 eq. did not change the reaction outcome. An alternative iodination protocol by Emmanuvel *et al*.[321] employing a $NaIO_4$/KI/NaCl mixture in acetic acid (AcOH) turned out inapplicable for this particular starting material and did not result in product formation, although ICl is supposed to form *in situ*.

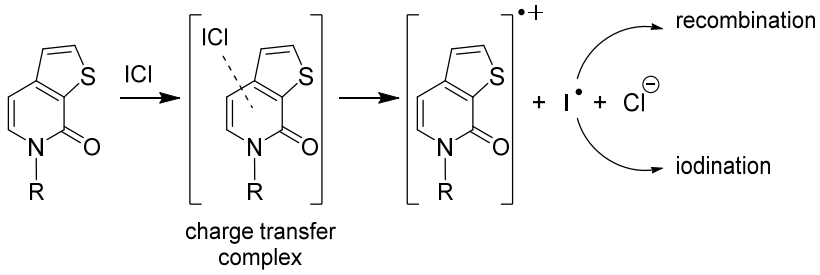
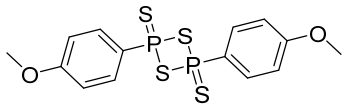


Figure 45. Schematic iodination mechanism with ICl according to Hubig *et al.*[320]; sugar residues (R = 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)) are omitted for clarity.

Further studies focusing on this topic could involve slow or iterative and possibly more dilute addition of ICl. This would suppress a hypothesized iodine radical recombination. Furthermore, reagent combinations like indium triflate ($In(OTf)_3$) with ICl[322] could help enhancing this reaction step and could be implemented in further studies.

Thionation[194,323,324]

The reaction of **49** with Lawesson's reagent (**53**, Figure 46) generally resulted in satisfying success. Yields could be optimized to 88 % from former 72 % by prolonged reaction times from 6.5 h to 11 h.



53

Figure 46. Chemical structure of Lawesson's reagent (**53**).

However, the thionation process is seemingly very slow, making it prone to negative influences like moisture. Conduction under inert atmosphere and exclusion of water turned out to be essential as in some cases the mixture turned black and neither product nor starting material could be isolated. Accurate Schlenk technique should be applied and the use of small batches of Lawesson's reagent (**53**) is recommended to minimize potential substance degradation over time. Proper pre-drying of starting material **49** by co-evaporation with anhydrous toluene as well as desiccation of thionating reagent **53**, *e.g.* by overnight submission to high vacuum, facilitate easy conversion of the amide oxygen atom to sulfur.

Deprotection of the ribose scaffold[313,314,318]

Removal of the ribose acyl protecting groups proceeded smoothly in overnight treatment of **50** with a methanolic ammonia solution. In this reaction acetamide is formed as byproduct which cannot easily be removed by drying under high vacuum since it has a boiling point of ~220 °C. Chromatographic purification of nucleoside **51** is thus inevitable to provide high-purity starting material for the following Sonogashira reaction.

Functionalizing the nucleobase analog[194]

Transition metal-catalyzed coupling of **51** to TMS-protected cyclopropene linker **36** is feasible under standard conditions[168,234,325,326] in typical yields of ca. 75 %. In principle, full conversion of the starting material can be observed by thin layer chromatography (TLC). However, due to the free hydroxyl groups and the resulting polarity of compound **52**, a certain loss of product evolves from incomplete elution during chromatographic purification. Future works involving the synthesis of r**TPT3^{CP}** TP (**34**) could thus consider modifying the current route. Cyclopropene linker **36** is potentially well-attachable to iodinated and thionated compound **50**. This species still comprises acyl protecting groups, making it easier to handle.

TMS deprotection of the methyl cyclopropene moiety[209,239,247]

TMS deprotection using TBAF gave r**TPT3^{CP}** nucleoside **35** in 77 % yield. Both crystalline TBAF trihydrate and a ready-to-use solution in THF were tested and resulted in similar reaction success. Occasionally, incomplete removal of TMS groups could be observed, adding generous amounts of fluoride donor did not overcome this issue. Stronger fluoride sources such as HF·py could possibly enable quantitative reactions; however, its hazard potential makes it a rather unattractive reagent.

At this stage the IEDDA cycloaddition properties were tested by reacting **35** with tetrazine-fluorophore conjugate **^{H}Tet-OG** (**28**, synthesized by C. Domnick[327], fluorophore: Oregon Green 488, Figure 47A). For this purpose, a threefold excess of the r**TPT3^{CP}** nucleoside (**35**) was incubated with **28** and turn-on fluorescence was tracked by iterative scanning measurements (λ_{ex} = 495 nm, Figure 47B). After only 20 min a plateau was reached (Figure 47C), indicating reaction completion. As control, iodinated nucleoside **51** was incubated with tetrazine-fluorophore conjugate **28** for 1 h and fluorescence scanned (Figure 47B, dashed line). Additionally, the reaction mixture was analyzed via high performance liquid chromatography coupled to mass spectrometry (HPLC-MS) and exhibited one major peak corresponding to the desired IEDDA click product **55** (Figure 47D and E).

These findings strongly support a pronounced and fast reactivity within the time frames suitable for in-cell applications in future experiments and most likely account for r**TPT3^{CP}** TP (**34**) and RNAs modified with this nucleotide in a similar fashion.

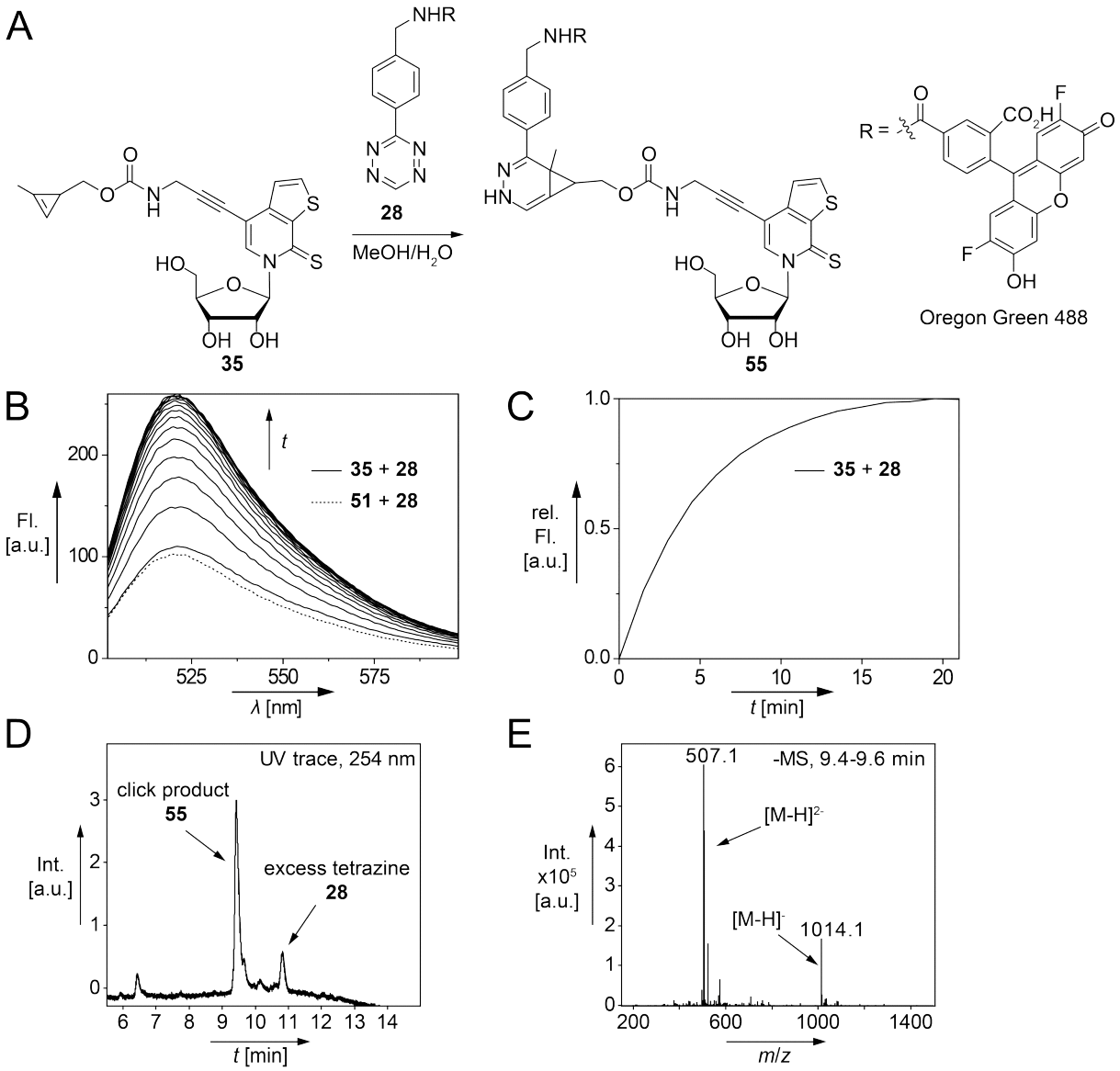


Figure 47. IEDDA cycloaddition of r**TPT3^{CP}** nucleoside **35** with tetrazine-fluorophore conjugate **28**; A. Reaction scheme; B. Reaction tracking by fluorescence scans (λ_{ex} = 495 nm); C. Reaction time course at λ_{em} = 495 nm; D. HPLC UV trace (5→100 % MeCN/0.1 % formic acid in 20 min, *Zorbax* column, see 5.2.4) of the reaction mixture; E. Electron spray ionization $(ESI)^-$ spectrum of the peak eluting at ~9.5 min, $M_{calcd.}$ = 1014.2 for $[M-H]^-$ of cycloaddition product **55**.

Triphosphate synthesis[130,187,188,315,328,329]

As final synthetic step, compound **35** was converted into r**TPT3^{CP}** triphosphate **34** applying Ludwig conditions[315]. This widespread method involves phosphorylation with $POCl_3$[330] in trimethylphosphate (Me_3PO_4) to give nucleoside-5′-phosphorodichloridates, which undergo further reaction with bis-tributylammonium pyrophosphate ($(Bu_3NH)_2PPi$) to 5′-cyclic triphosphates (Figure 48). Neutral hydrolysis yields the desired nucleoside-5′-triphosphates.

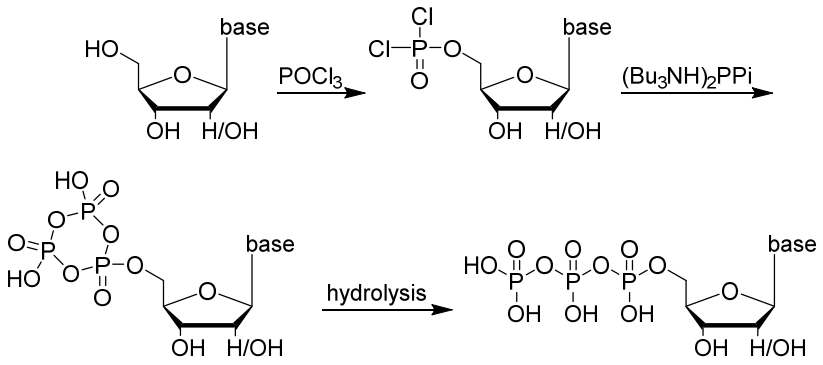


Figure 48. Schematic generation of nucleoside 5′-triphosphates applying the Ludwig method.[315,330]

With an isolated yield of 31 % after preparative high performance liquid chromatography (HPLC) purification, satisfactory amounts of pure r**TPT3^{CP}** triphosphate **34** could be obtained. Purity and integrity were confirmed by HPLC-MS analysis (Figure 49).

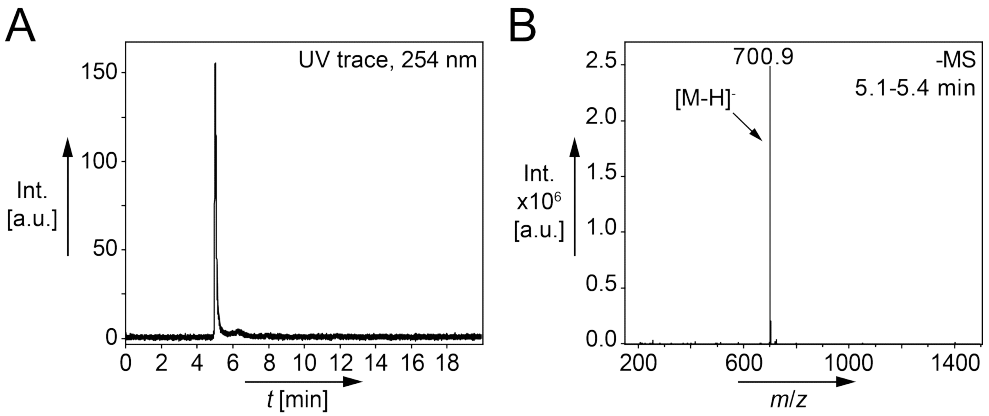


Figure 49. HPLC-MS analysis of purified r**TPT3^{CP}** TP (**34**) (5→100 % MeCN/0.1 % (*w*/*v*) ammonium acetate (NH_4OAc) in 20 min, *Zorbax* column, see 5.2.4); A. UV trace at 254 nm, B. ESI$^-$ spectrum of **34** eluting at ~5.2 min ($M_{calcd.}$ = 701.0 for $[M-H]^-$).

This particularly sensitive cascade was conducted under complete exclusion of moisture (and oxygen) as the highly reactive phosphorodichloridate intermediate is prone to side reactions. Low amounts of starting material available in a research laboratory after many challenging and complex synthetic steps (usually 20-100 mg) further complicate the delicate handling. Even with the highest possible caution, triphosphate syntheses failed in some cases and desired triphosphate compounds could not be detected during HPLC-MS analysis of crude reaction mixtures. A major bottleneck of this one-pot reaction seems to be the successful pyrophosphate addition as significant amounts of 5′-monophosphate are often observed analyzing crude reactions (Figure 50).

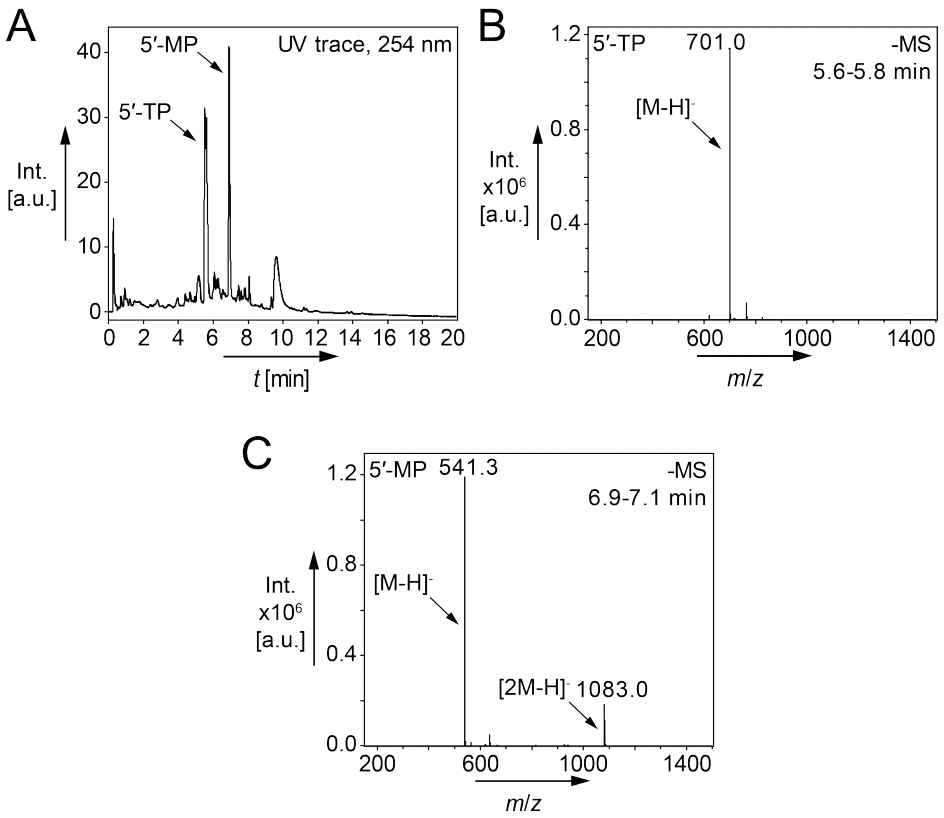


Figure 50. HPLC-MS analysis of a crude r**TPT3^{CP}** TP (**34**) synthesis (5→100 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, *Zorbax* column, see 5.2.4); A. UV trace indicating triphosphate product (**34**, 5′-TP) and the major monophosphate byproduct (5′-MP); B. ESI^- spectrum of the peak eluting at ~5.6 min, corresponding to **34** ($M_{calcd.}$ = 701.0 for $[M-H]^-$); C. ESI^- spectrum of the peak eluting at ~7 min, corresponding to r**TPT3^{CP}** 5′-MP ($M_{calcd.}$ = 541.1. for $[M-H]^-$).

In this study, standard C_{18} reversed-phase HPLC purification was employed and usually gave satisfactory results. However, in several attempts purification of the complex mixtures proved to be challenging and iterative HPLC purification had to be applied.

Due to the high polarity of nucleoside-5′-triphosphates, regular-phase HPLC might be advantageous. Literature procedures frequently employ ion exchange chromatography prior HPLC purification[130,168,196,305,331], which might also help overcome separation issues. The Romesberg group recently stated that in their hands most reproducible results on unnatural deoxyribonucleoside triphosphates were obtained using anion exchange chromatography on a diethylaminoethyl(DEAE)-*Sephadex* matrix.[332]

Despite early syntheses of naturally occurring r**A** TP and d**A** TP were reported to yield 78 % and 86 % of isolated product[315], respectively, the Ludwig approach still seems to be challenging as current attempts generating unnatural nucleotides generally result in ca. 5–40 % of the desired (deoxy)nucleoside 5′-triphosphates[168,194,305,333].

An alternative approach for the generation of nucleoside-5′-triphosphates was developed by Ludwig and Eckstein[334], employing salicyl phosphorochloridite (**56**, Figure 51) instead of $POCl_3$ in the first reaction step. This protocol is reported to result in formation of fewer

byproducts[328] and could thus increase isolable yields in future attempts. However, protection of the 3′-hydroxyl group is required, adding further steps to the overall synthetic route.

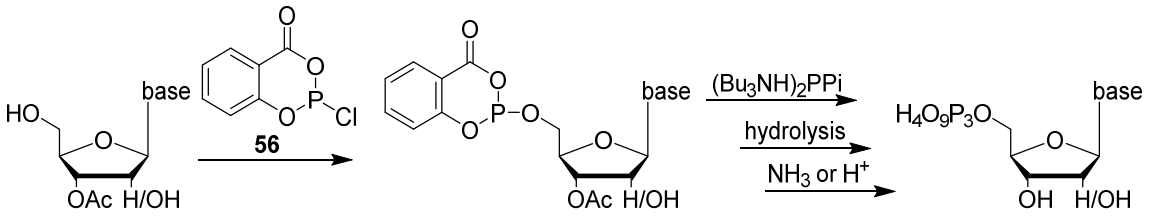


Figure 51. Schematic representation of the Ludwig-Eckstein triphosphate synthesis.[328,334]

In summary, the presented synthetic route facilitates generation of the novel ribose triphosphate **34** containing a cyclopropene-functionalized derivative of the unnatural base **TPT3**. Several steps of the cascade were improved compared to former[298] works, further enhancing availability and cost efficiency of r**TPT3^{CP}** TP (**34**) production. The cyclopropene linker attached to the nucleobase analog enables additive-free IEDDA cycloadditions with tetrazines bearing reporter groups.

3.1.2 Applications of r**TPT3^{CP}** TP (**34**) in site-specific RNA labeling

Transcription reactions from d**NaM**-containing DNA templates in the presence of r**TPT3^{CP}** TP (**34**) were envisaged to allow for facile and efficient preparation of site-specifically modified RNA molecules in significant quantities. Several approaches for predefined r**TPT3^{CP}** labeling and posttranscriptional modification of RNAs have been developed during this study.

Short templates served for direct T7 *in vitro* transcription, which enabled site-specific modification of ribonucleotide oligomers with an upper limit of approx. 160 nt as shown in section 3.1.2.1.

Preparation of r**TPT3^{CP}**-modified RNA sequences with increased length is described in section 3.1.2.2 Transcription from dsDNA templates assembled via a fusion PCR approach extended the applicability of the labeling approach to constructs of up to 500 nt.

In vitro transcription from a UBP-containing plasmid DNA is presented in section 3.1.2.3, which sets the base for RNA site-specific functionalization potentially without length limitation.

Scope and limitations of these new techniques are presented and carefully revised in the following.

3.1.2.1 Labeling of short RNAs (up to ca. 160 nt)

To assess the fundamental feasibility applying r**TPT3^{CP}** TP (**34**) for template-directed incorporation into RNA via *in vitro* transcription from d**NaM**-modified DNAs, several short target ribonucleic acid oligonucleotides were generated and systematically evaluated.

Production of an r**TPT3^{CP}**-modified reference RNA construct (**RNA^{CP}**)

Initially, a 35 nt DNA, containing one d**NaM** nucleotide (**DNA^{NaM}**), was used in transcription reactions. Spanning the T7 class III promoter region[48,335], the sequence could be hybridized to a primer strand (**Primer^{n1}**) enabling the T7 RNA polymerase to initiate ribooligonucleotide production. The template sequence **DNA^{NaM}** has been successfully employed in earlier studies on transcription properties of the unnatural base pair (UBP) **5SICS**:**NaM** by the Romesberg group.[187,190] Therefore, it served as an excellent reference construct investigating the incorporation of r**TPT3^{CP}** nucleotides into RNA.

DNA^{NaM} was transcribed into the corresponding 17 nt **RNA^{CP}** (Figure 52A) employing r**TPT3^{CP}** TP (**34**) in a standard T7 *in vitro* transcription protocol as established internally (see 5.3.6). This approach relies on the novel UBP combination **TPT3^{CP}**:**NaM** (Figure 52B). HPLC-MS analysis of the resulting crude transcript revealed a surprisingly homogenous reaction outcome as can be seen from the UV trace (Figure 52C, left panel). The major peak of the obtained spectrum was assigned to the 5′-triphosphate (5′-TP) product of **RNA^{CP}** as sodium adducts (Figure 52C, right panel).

Pronounced ability of T7 RNA polymerase to accept r**TPT3^{CP}** TP (**34**) as substrate and to incorporate these unnatural nucleotides into RNA in a template-directed manner from suitable d**NaM**-containing DNA could thus impressively be shown.

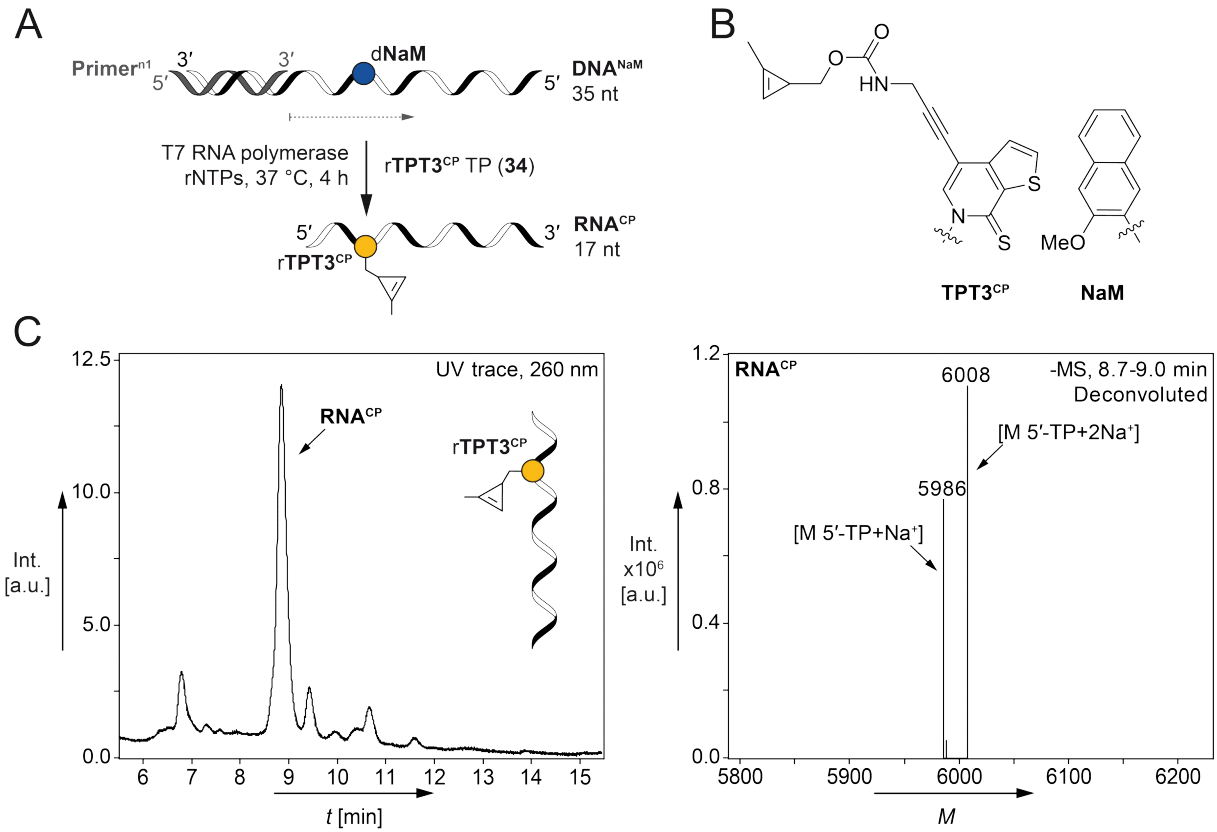


Figure 52. Site-specific r**TPT3^{CP}**-labeling of a short RNA; A. Scheme for template-directed incorporation of an r**TPT3^{CP}** nucleotide, producing **RNA^{CP}** via *in vitro* transcription from **DNA^{NaM}** (enzymatic synthesis direction indicated by grey dashed arrow); B. The novel UBP **TPT3^{CP}**:**NaM** used in this study; C. HPLC-MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) of a crude, DNase-digested, and gel-filtrated transcript sample: UV-trace (left panel) and deconvoluted ESI$^-$ mass spectrum of the peak eluting at ~8.8 min (right panel, $M_{calcd.}$ = 5985 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 6007 for [M 5′-TP+2Na$^+$]).

Next, fluorophore attachment via IEDDA click reaction of *in vitro* prepared **RNA^{CP}** with a tetrazine-Oregon Green *488* conjugate (**^{H}Tet-OG**, **28**, synthesized by C. Domnick[232,234,327,336]) was performed, which allowed for sensitive analysis of post-transcriptionally fluorophore-labeled **RNA^{CP}-Tet-OG** (Figure 53A).

Gel electrophoretically purified **RNA^{CP}** was subjected to overnight incubation with **28** (7-fold excess) at room temperature. The UV trace of an HPLC-MS analysis of the reaction revealed presence of click product **RNA^{CP}-Tet-OG** as major component as well as residual unreacted **RNA^{CP}** and leftover tetrazine **28** (Figure 53B). Incomplete reaction is expected as common procedures involve large tetrazine excess of 50 eq.[331] up to several hundred molecular equivalents[305] to ensure full conversion. A double peak appearing for the fluorophore-labeled RNA presumably results from structural isomers of the IEDDA cycloaddition which have also been observed performing similar experiments on r**U^{NOR}**- and r**TPT3^{NOR}**-modified RNA[234,337].

Both the identities of **RNA^{CP}** and click product **RNA^{CP}-Tet-OG** could be confirmed by mass spectrometric analysis (Figure 53C and D, respectively) and were detected as 5′-monophosphate (5′-MP), 5′-diphosphate (5′-DP), and 5′-triphosphate (5′-TP) species. Technically, *in vitro* transcribed RNAs contain a 5′-terminal TP, however, partial hydrolysis to 5′-DP and 5′-MP molecules was commonly observed after purification procedures.

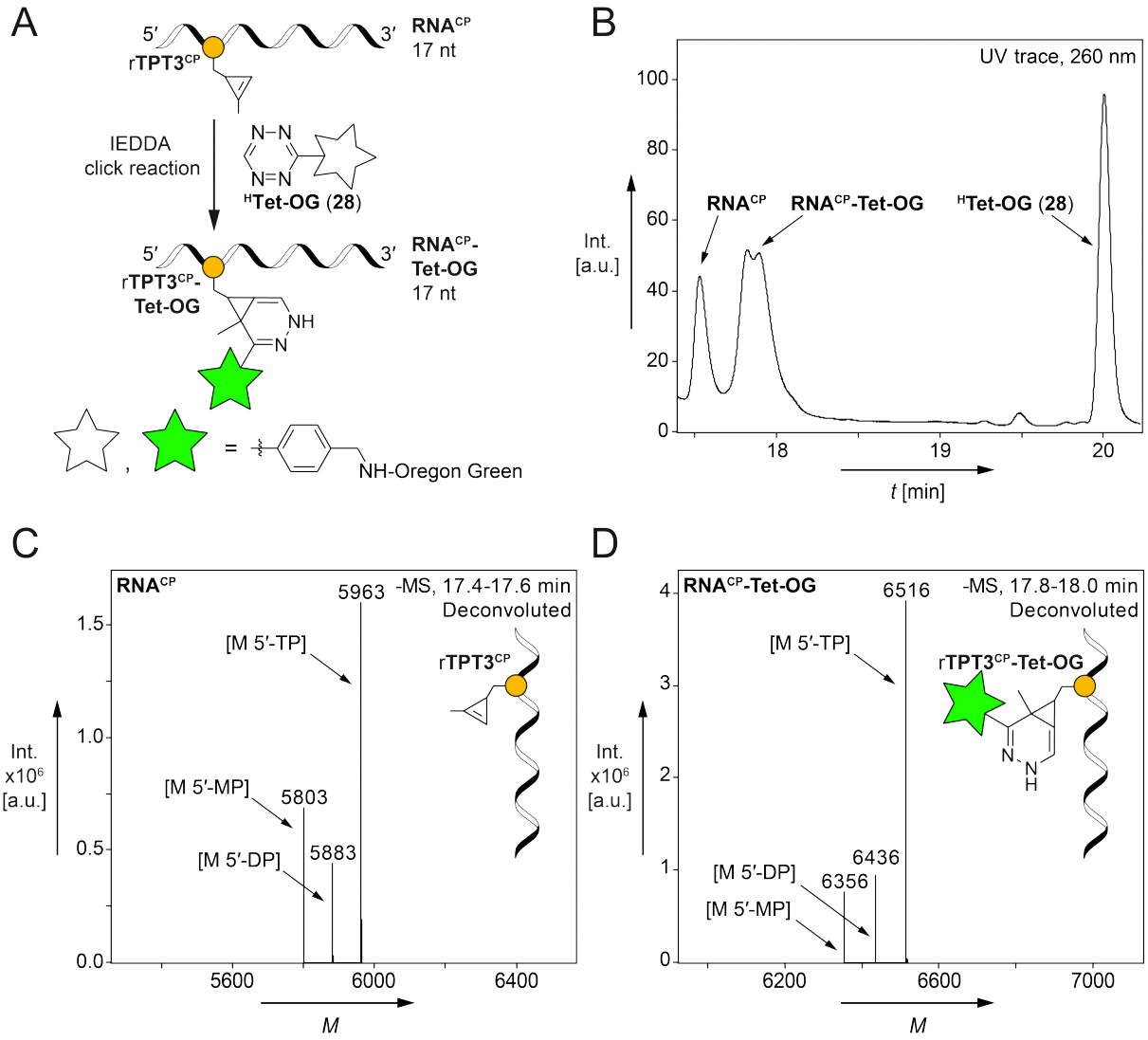


Figure 53. IEDDA reaction of **RNA^{CP}** with **^{H}Tet-OG** (**28**); A. Schematic representation; B.-D. HPLC-MS analysis of the click reaction (3 % MeCN for 9 min followed by a gradient 3→40 % MeCN/buffer in 11 min, 0.2 mL min^{-1}, column *Xterra*, see 5.3.13): B. UV trace of an overnight reaction of **RNA^{CP}** with 7 eq. **28**; C. Deconvoluted ESI-MS spectrum of **RNA^{CP}** eluting at ~17.5 min ($M_{calcd.}$ = 5802 for [M 5′-MP], $M_{calcd.}$ = 5882 for [M 5′-DP], $M_{calcd.}$ = 5962 for [M 5′-TP]); D. Deconvoluted ESI-MS spectrum of **RNA^{CP}-Tet-OG** eluting at ~17.9 min ($M_{calcd.}$ = 6356 for [M 5′-MP], $M_{calcd.}$ = 6436 for [M 5′-DP], $M_{calcd.}$ = 6516 for [M 5′-TP].

As further proof for successful and specific r**TPT3^{CP}** incorporation into RNA and subsequent IEDDA click application on modified sequences, DPAGE was performed comparing crude, DNase-digested **RNA^{CP}** with the purified construct (Figure 54). As control, **RNA^{n1}** was transcribed from the unmodified DNA template with only canonical nucleotides added to the

reaction mix. Each construct was analyzed before and after 1 h incubation with 1 eq. tetrazine **28**. After DPAGE separation, the gel was first fluorescence scanned and then submitted to *SYBR Safe* nucleic acid post-staining. Turn-on fluorescence upon IEDDA reaction could easily be detected as intense fluorescent band of **RNA^{CP}-Tet-OG** in both crude and purified transcripts. By contrast, the absence of such in the case of **RNA^{n1}** incubated with **28** reassured the fluorescence signal to specifically originating from template-directed r**TPT3^{CP}** nucleotide incorporation into RNA and successful click reaction on the modified nucleic acid. Interestingly, a faint ladder of bands with faster migration behavior than **RNA^{CP}-Tet-OG** is visible in the fluorescence scan. For such short constructs near the end of initiation phase (≤12 nt[48,338]), it is arguable that T7 RNA polymerase falls off the template when approaching the unnatural base to some extent. Since the d**NaM** modification in the template is also positioned four nucleotides from the 5′-end the signals are likely to correspond to a -1, -2, and -3 transcript relative to the full-length RNA. These bands are not visible in unspecific nucleic acid stain and are thus estimated to appear in a negligible range (the lower detection limit of *SYBR Safe* is 0.5 ng per band for constructs ≥200 nt, *thermofisher.com*, requested 2019/08).

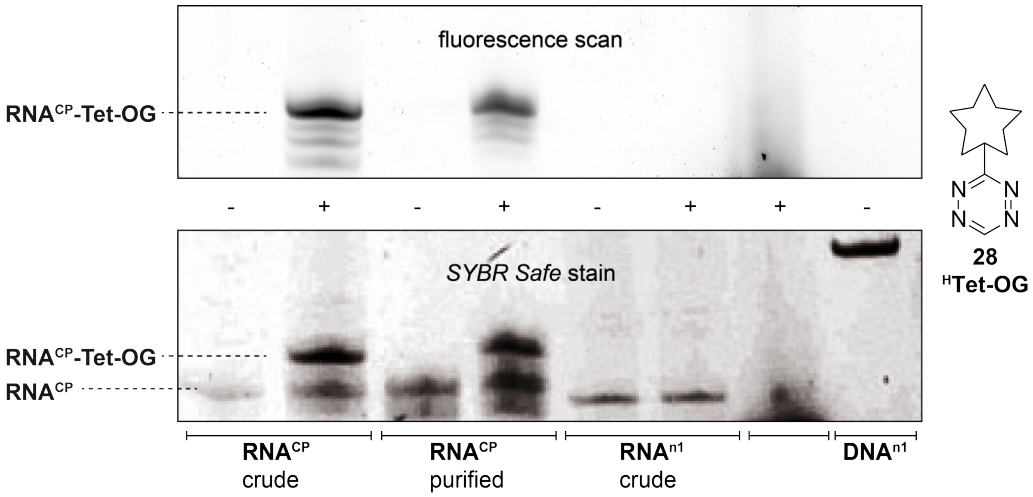


Figure 54. 20 % DPAGE analysis of **RNA^{CP}** transcript and IEDDA click reaction with 1 eq. tetrazine **28** giving **RNA^{CP}-Tet-OG** in fluorescence scanning mode (upper panel) and *SYBR Safe* post-stain (lower panel).

These data prove the successful and efficient introduction of r**TPT3^{CP}** TP (**34**) into RNA during transcription reaction from d**NaM**-containing **DNA^{NaM}**. The straight-forward experimental design is well-applicable even to crude reactions and facilitates post-transcriptional decoration of RNA with reporter groups under particularly mild IEDDA cycloaddition conditions.

Fluorescently labeling a tRNA sequence at its anticodon loop (**tRNA^{CP}**)

In order to investigate versatility and applicability preparing longer RNA sequences of defined tertiary structure, the approach was employed to site-specifically label the *Methanocaldococcus jannaschii* $tRNA_{CUA}^{Tyr}$ (77 nt) with an r**TPT3^{CP}** nucleotide placed in its anticodon loop.

A template DNA (**tDNA^{NaM}**) was hybridized with its complementary sequence (**tDNA^{RV}**) causing a d**NaM**:d**T** mismatch (in analogy to the schematic procedure depicted in Figure 52A). Transcription efficiency turned out to be more efficient when constructs >50 nt were transcribed from fully double-stranded templates. Mismatches of d**NaM** nucleotides and canonical bases were tolerated in the tested constructs, reverse strands could thus be synthesized containing only native nucleotides. Notably, proper hybridization of DNA sequences might be restricted in length; PCR-based approaches are favorable preparing long dsDNA templates for transcription (see sections 3.1.2.2 and 3.1.2.3).

Transcription in the presence of r**TPT3^{CP}** TP (**34**) enabled preparation of the potentially highly structured RNA construct **tRNA^{CP}** (Figure 55A). Subsequent IEDDA click reaction of the crude RNA with tetrazine **28** resulted in efficient fluorescent labeling (**tRNA^{CP}-Tet-OG**), which was analyzed by DPAGE (Figure 55B) since reasonable detection via HPLC-MS is not easily possible for transcribed RNA of extended length. The click product **tRNA^{CP}-Tet-OG** could be detected as strong fluorescent band, whereas the control **tRNA^{n}** (transcribed from the corresponding native DNA template **tDNA^{n}**) did not give any signal when incubated with tetrazine **28**.

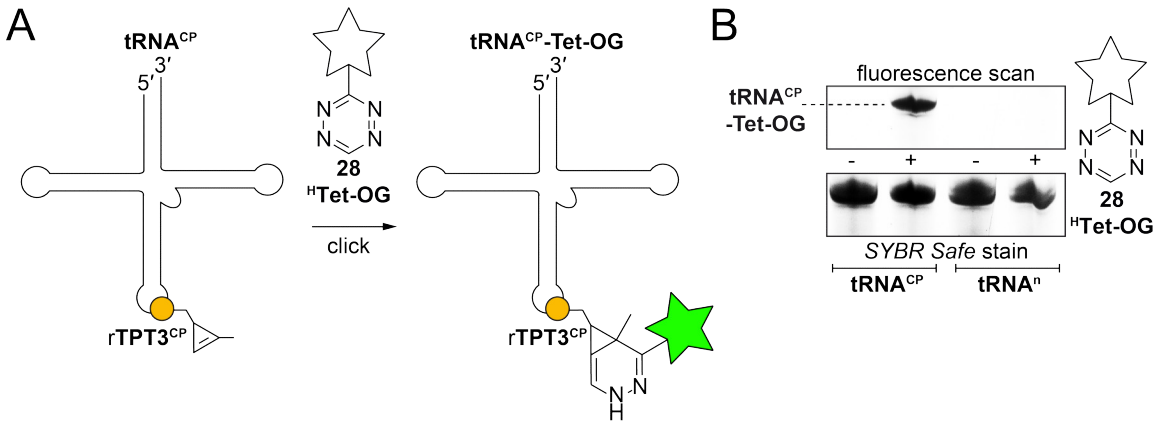


Figure 55. Site-specific labeling of a tRNA sequence; A. Schematic representation of **tRNA^{CP}** reacting with tetrazine-fluorophore conjugate **28** to **tRNA^{CP}-Tet-OG**; B. Comparative 20 % DPAGE analysis of **tRNA^{CP}**, clicked **tRNA^{CP}-Tet-OG**, and unmodified **tRNA^{n}** by fluorescence scanning (upper panel) and *SYBR Safe* post-stain (lower panel).

The new method to site-specifically label RNA with cyclopropene-modified r**TPT3^{CP}** nucleotides by *in vitro* transcription and visualization by subsequent reaction with tetrazine-fluorophore conjugates is hereby shown to be applicable to longer constructs of complex tertiary structure.

With **tRNA^{CP}**, a sensitively folded site-specifically modified RNA was prepared. However, one could argue that the approach lacks definitive evidence of the precise structure being developed. It is known, that *in vitro* transcribed (and purified) RNA is in many cases misfolded.[339–342] With the unnatural r**TPT3^{CP}** nucleotide introduced into RNA, chances of misfolding might increase. Production of functional RNAs such as self-cleaving ribozymes could serve to analyze the r**TPT3^{CP}** modification effects via analysis of their catalytic activity and was therefore attempted next.

Site-specific r**TPT3^{CP}**-labeling of the human *CPEB3* ribozyme by K. Kulikov[343,344]

r**TPT3^{CP}** RNA labeling was also applied to characterization studies of the human HDV-like *CPEB3* ribozyme.[343] *CPEB3* self-cleavage is Mg^{2+}-dependent. T7 *in vitro* transcription, however, needs high amounts of Mg^{2+} ions (15–30 mM).[48] Hence the main reactive species of *CPEB3* ribozyme is not accessible by *in vitro* prepared RNA due to premature cleavage. A radioactively labeled *trans*-cleaving ribozyme variant unfortunately gave poor results in terms of unexpectedly low cleavage activity. Most likely, the two-strand technique hindered complex and cleavage-essential secondary structure formation.[345]

r**TPT3^{CP}** labeling of the target *CPEB3* RNA sequence allowed performing a co-transcriptional assay (experiments by K. Kulikov), enhancing the possibility of correct ribozyme folding and thus giving most plausible experimental results. A single r**TPT3^{CP}** modification near the 5′-end of the ribozyme (**CPEB3^{CP}**, Figure 56A) advantageously enabled direct correlation of the full ribozyme and its split product via fluorescent readout. During a co-transcriptional cleavage assay, **CPEB3^{CP}** samples were labeled with a tetrazine-BODIPY conjugate (**^{H}Tet-BDP**, **57**, Figure 56B) to sensitively detect **CPEB3^{CP}-Tet-BDP** and its cleaved fragment **CPEB3^{CP}-2-Tet-BDP** based on fluorescence readout after DPAGE separation (Figure 56C).

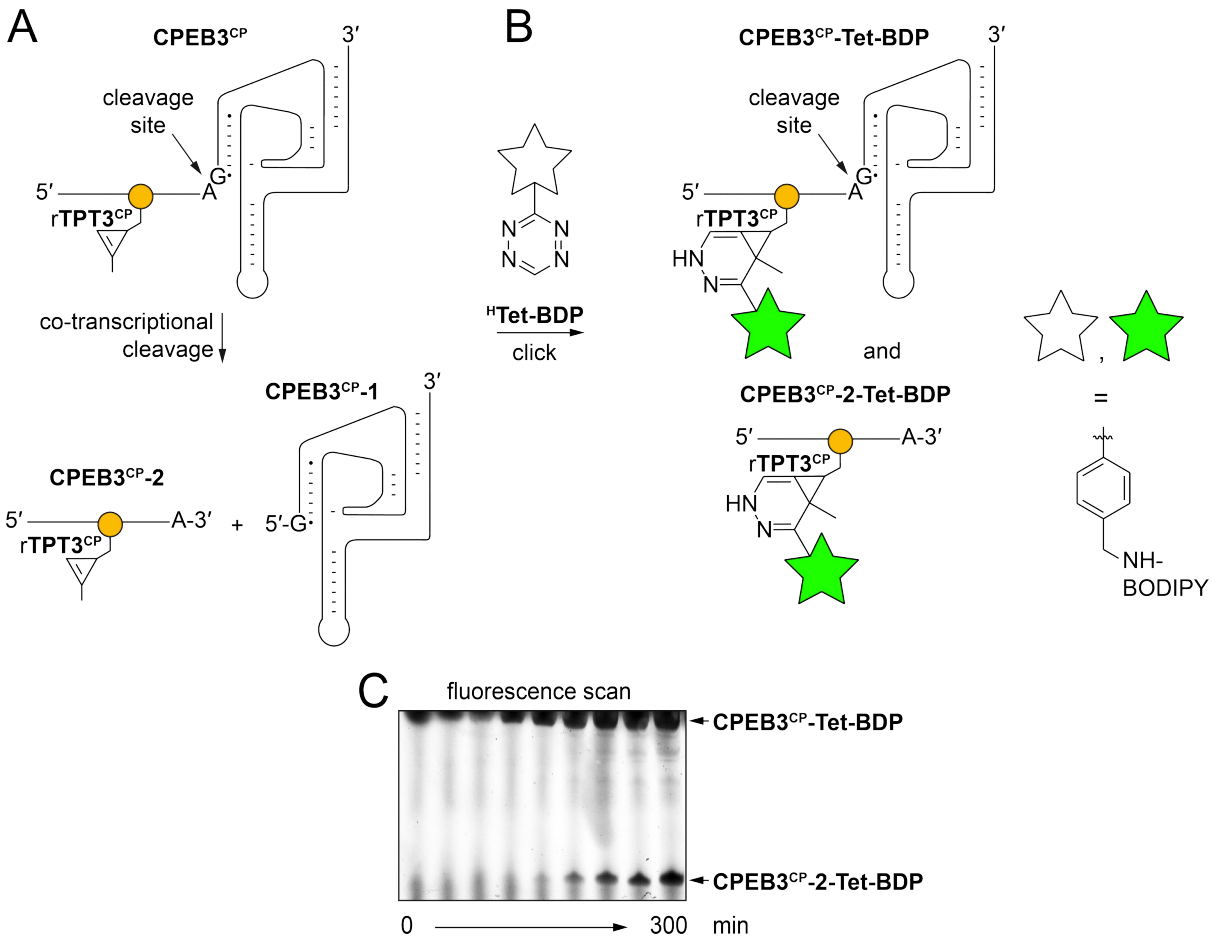


Figure 56. Site-specific r**TPT3^{CP}**-labeling of the *CBEB3* ribozyme; A. Schematic representation of **CPEB3^{CP}** and its cleavage process; B. IEDDA click reaction of r**TPT3^{CP}**-modified *CPEB3* constructs with tetrazine-BODIPY conjugate **57**; C. Co-transcriptional self-cleavage assay, data adapted from K. Kulikov[40],[41]: fluorescence scan of a 20 % DPAGE analysis showing **CPEB3^{CP}-Tet-BDP** and the cleaved product **CPEB3^{CP}-2-Tet-BDP**.

Evaluation of the reaction profile (not shown) gave impressively consistent data compared to values reported by Bevilacqua and co-workers who previously established a similar assay based on ^{32}P-labeling of the ribozyme[346].

K. Kulikov's results prove functionality and thereby proper folding of the *in vitro* transcribed **CPEB3^{CP}** RNA, emphasizing the versatility of the r**TPT3^{CP}** labeling approach even to investigate complex ribozyme kinetics. Furthermore, the presented assay can relieve researchers from potentially harmful radioactive experiments. In times of high regulation of radioactivity this is a highly favorable feature for future benefit.

3.1.2.2 Labeling of mid-range RNAs (up to approx. 450 nt)

Successful preparation of a functional *in vitro* transcribed ribozyme containing the unnatural r**TPT3^{CP}** nucleotide was demonstrated. However, the presented approach is limited to restrictions of cost-effective solid-phase DNA synthesis (≤200 nt). Therefore, further advances were necessary to expand the toolbox of site-specific RNA modifications using an expanded genetic alphabet allowing prospective biochemical studies on desired ncRNA molecules.

Target RNA: the *Bacillus subtilis glmS* ribozyme

A 185 nt sequence (position -27 to +159 according to Breaker and co-workers[61]) representing the catalytically active part of the *B. subtilis glmS* ribozyme[46–49] with a 27 nt 5′-terminus was chosen as a new model system to be prepared via *in vitro* transcription. Two unnatural modifications were envisaged, placing r**TPT3^{CP}** nucleotides in the loop region of helix P1 and P4a, respectively (Figure 57).

As template DNA, a corresponding 202 nt d**NaM**-modified template DNA including the T7 promoter region was required. This construct exceeds the length of reasonable chemical solid-phase synthesis; therefore, a fusion PCR-based approach preparing the respective double stranded DNA (dsDNA) was established.

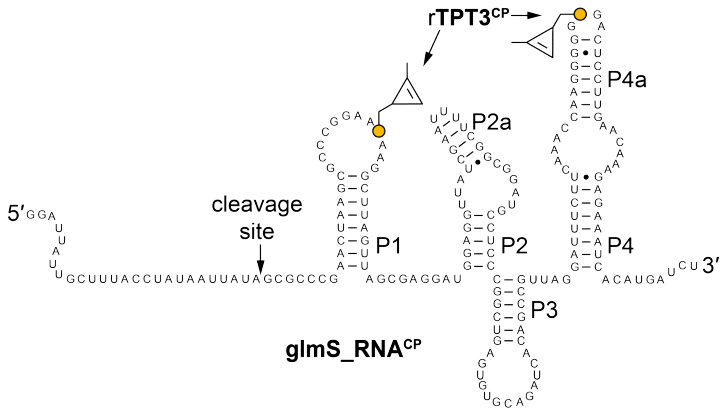


Figure 57. Proposed r**TPT3^{CP}**-modified *glmS* ribozyme construct **glmS_RNA^{CP}**; original sequence was adapted from Winkler *et al.*[61], r**TPT3^{CP}** nucleotides substituting for r**A** and r**U** in the loops of helices P1 and P4a, respectively.

Assembly of the d**NaM**-modified full-length dsDNA template using fusion PCR

For preparation of doubly r**TPT3^{CP}**-modified **glmS_RNA^{CP}** via T7 *in vitro* transcription, a two-step PCR strategy towards the template DNA by Shevchuk *et al.*[348] was adapted to involve UBPs. Build-up and amplification of the full-length dsDNA was realized starting from three overlapping DNA oligonucleotides.

Besides canonical triphosphates the UBP triphosphates d**TPT3** TP (**22**) and d**NaM** TP[194] (**24**) were employed for this purpose. These deoxynucleotides were synthesized in-house as described in the literature either from scratch (d**TPT3** TP[194], **22**) or from the commercially available nucleoside (d**NaM** TP[188], **24**) both during this thesis (see section 3.4.1) and as part of the thesis work by C. Domnick[327].

The first step assembling the dsDNA template contained two anti-sense strands, **glmS_DNA^{1NaM}** and **glmS_DNA^{3NaM}**, as well as a splint sense sequence **glmS_DNA^{2n}** (Figure 58A). Within 5 cycles at an optimized annealing temperature (see section 5.3.4) *Taq* DNA polymerase was employed to give extended strands **glmS_DNA^{ext_sense}** and **glmS_DNA^{ext_anti}**. In parallel, the approach was performed with templates **glmS_DNA^{1n}** and **glmS_DNA^{3n}** consisting of canonical bases in absence of **22** and **24**.

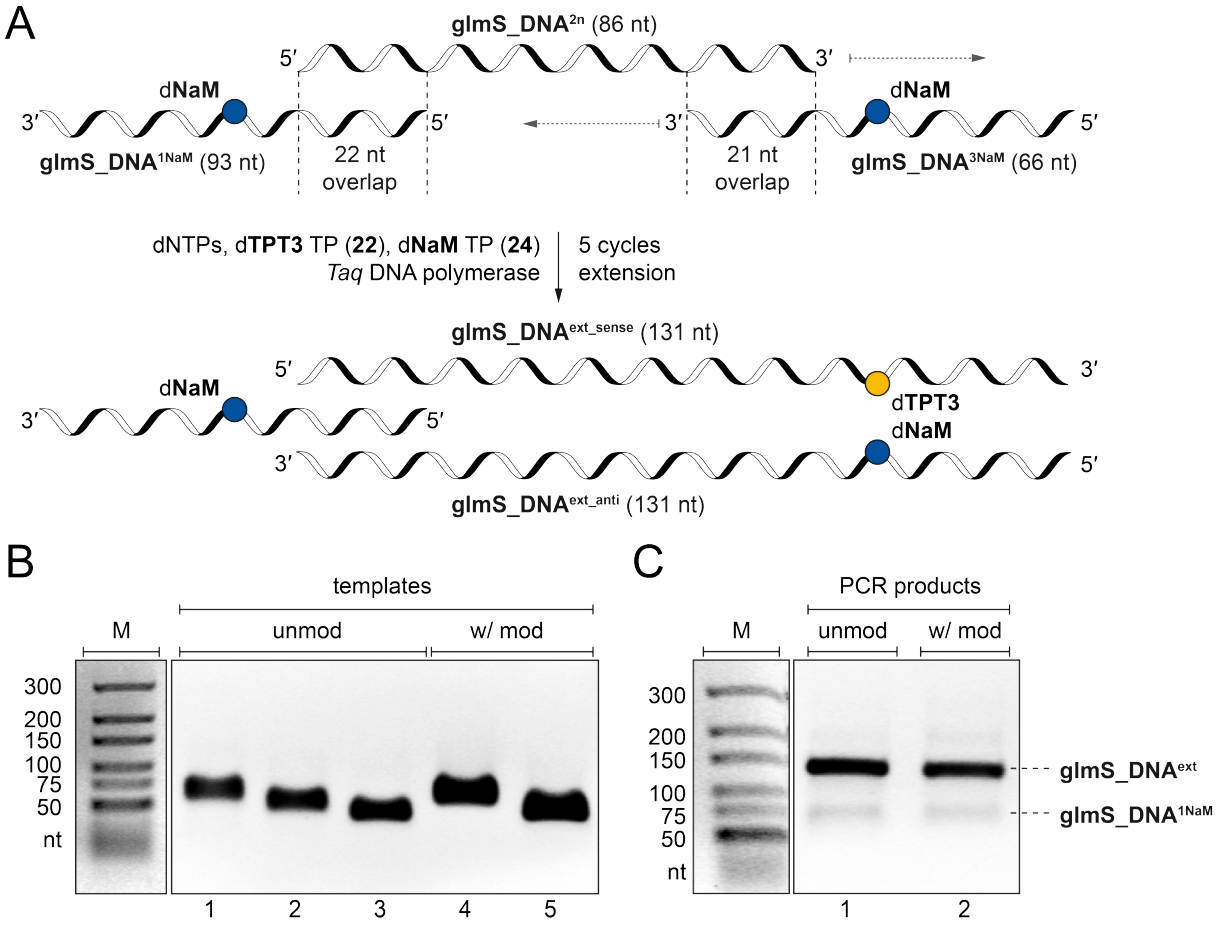


Figure 58. First step of dsDNA template generation for *in vitro* transcription of the doubly r**TPT3^{CP}**-modified ribozyme **glmS_RNA^{CP}**; A. Schematic overview of the extension approach with indicated overlap regions of the three template sequences; grey dashed arrows show the direction of enzymatic DNA synthesis; B. 2 % agarose gel electrophoretic analysis (EtBr stain) of unmodified and d**NaM**-modified DNA templates: M: Marker (*Ultra Low Range DNA Ladder*); Lane 1: **glmS_DNA^{1n}**, Lane 2: **glmS_DNA^{2n}**; Lane 3: **glmS_DNA^{3n}**; Lane 4: **glmS_DNA^{1NaM}**; Lane 5: **glmS_DNA^{3NaM}**; C. 2 % agarose gel (EtBr stain); M: Marker; Lane 1: Extension approach with unmodified templates and canonical dNTPs; Lane 2: Extension with d**NaM**-modified templates and d**TPT3** (**22**)/d**NaM** TP (**24**).

Notably, **glmS_DNA^{1NaM}** is not actively taking place in the extension approach and could thus be excluded from the corresponding reaction mix. This setup yet gave excellent experimental results and no need emerged to change the literature-based protocol[348].

The first step reaction mixture was then submitted to an amplification PCR including forward and reverse primers **glmS_Primer_FW^{1}** and **glmS_Primer_RV^{1}** (Figure 59A). Within 20 cycles of denaturation, annealing, and extension (see 5.3.4 for details) the full-length dsDNA construct could be generated and replicated, which was confirmed by agarose gel electrophoresis (Figure 59C).

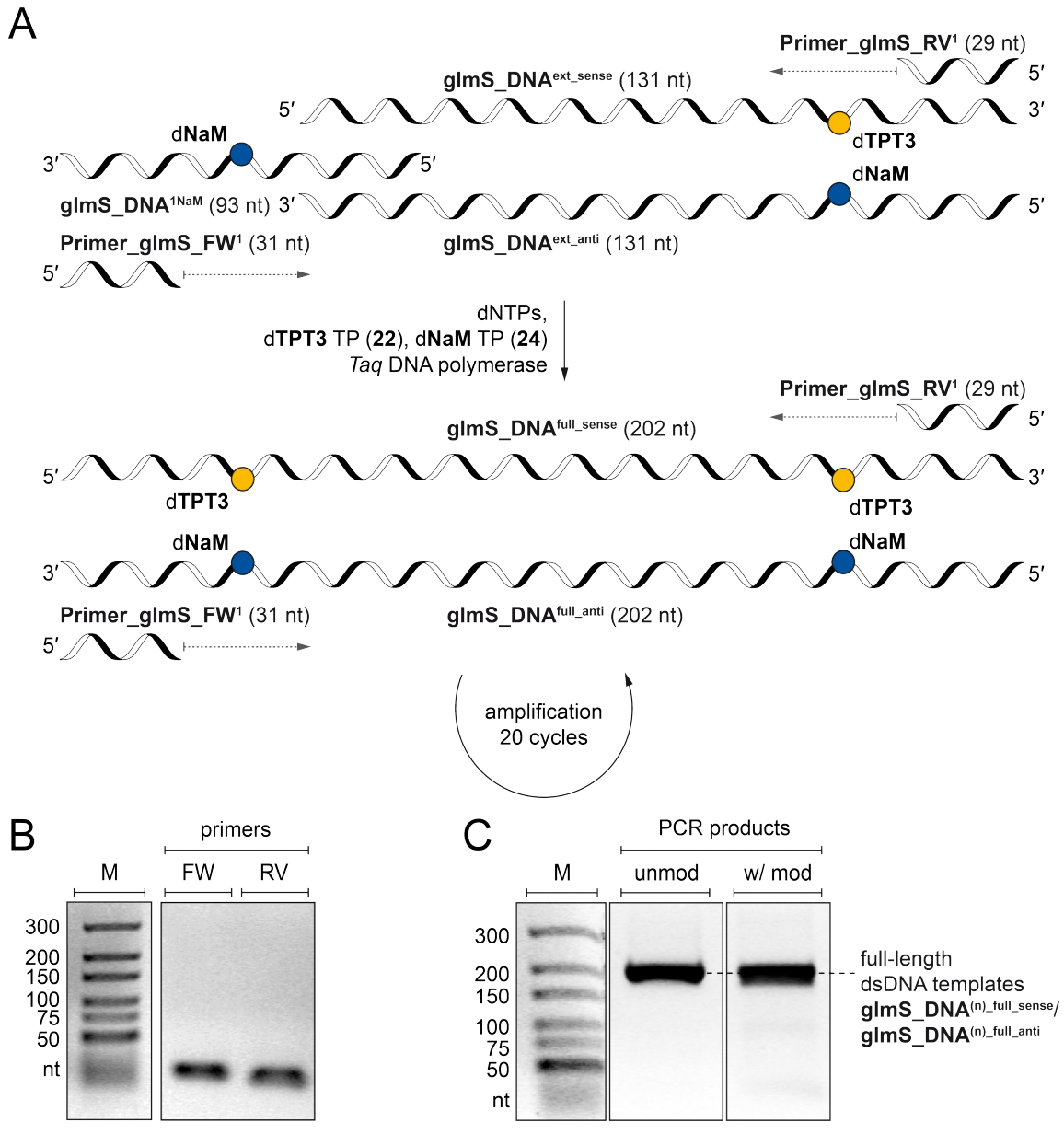


Figure 59. Second step of the fusion PCR approach preparing and amplifying the dsDNA template for *in vitro* transcription of **glmS_RNA^{CP}**; A. Schematic approach; grey dashed arrows indicate direction of DNA synthesis; B. 2 % agarose gel (EtBr stain) of the primers employed in the amplification PCR; C. 2 % agarose gel detection (EtBr stain) of target sequences in comparison with unmodified sequences.

Using this approach, it was possible to generate a 202 nt dsDNA template site-specifically modified with two d**NaM**:d**TPT3** base pairs. As can be seen from the qualitatively identically intense DNA bands on the agarose gel, the d**NaM**/d**TPT3**-modified PCR seems to proceed just as efficient as amplification of the native sequence. This confirms Romesberg's findings on the replication properties of the d**NaM**:d**TPT3** base pair[194].

The assembled full-length **glmS_DNA^{full}** sequences also comprised the T7 promoter region, enabling T7 RNA polymerase to initiate transcription. The DNA would thus facilitate template-directed incorporation of r**TPT3** nucleotides into RNA when suitable ribose triphosphates are provided to the reaction mix in the same manner as shown for the short, chemically synthesized template DNAs containing d**NaM** nucleotides (section 3.1.2.1).

In vitro preparation of a site-specifically fluorophore-labeled *glmS* ribozyme

Standard T7 *in vitro* transcription from **glmS_DNA^{full_anti}** in the presence of r**TPT3^{CP}** TP (**34**) was performed yielding **glmS_RNA^{CP}** as proposed above. The novel ribozyme variant site-specifically decorated with cyclopropene residues could then be visualized by submission to an IEDDA click reaction with tetrazine-fluorophore conjugate **^{H}Tet-BDP** (**57**) to give **glmS_RNA^{CP}-Tet-BDP** (Figure 60A). DPAGE analysis of the clicked and unclicked *glmS* variants clearly shows intense fluorescent bands for **glmS_RNA^{CP}-Tet-BDP** (Figure 60B, right panels), and the absence of such in the case of **glmS_RNA^{CP}** (Figure 60B, left panels) states successful incorporation of clickable r**TPT3^{CP}** moieties into the RNA of interest.

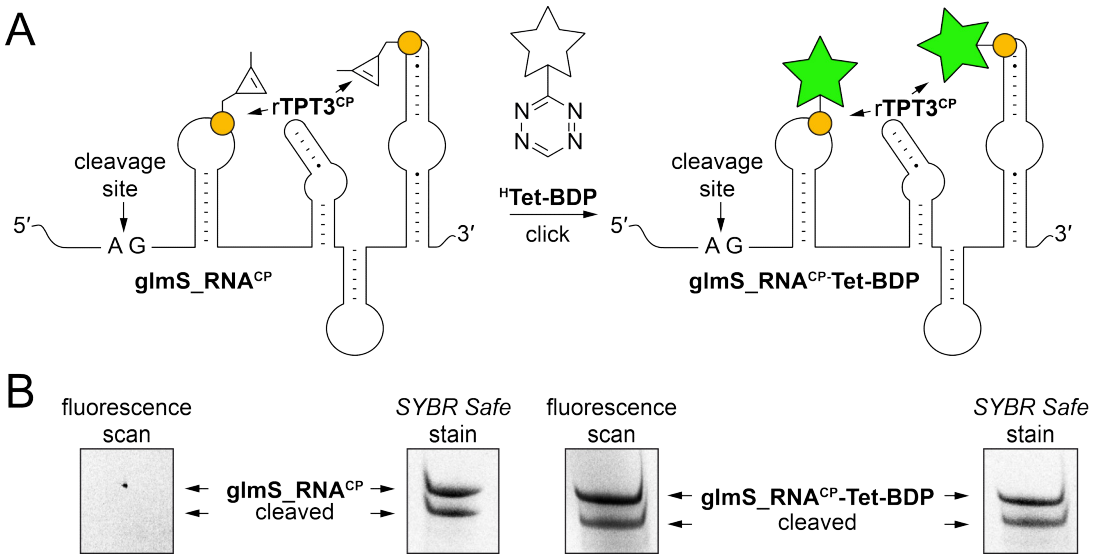


Figure 60. Site-specifically r**TPT3^{CP}**-labeled **glmS_RNA^{CP}** reacting with tetrazine **57** to yield the doubly fluorophore-tagged **glmS_RNA^{CP}-Tet-BDP**; A. Reaction scheme; B. 12 % DPAGE analysis of unreacted **glmS_RNA^{CP}** and **glmS_RNA^{CP}-Tet-BDP** comparing fluorescence scan and *SYBR Safe* staining of the same gel.

'*Click and cleave*': Assessing the self-cleavage kinetics of **glmS_RNA^{CP}-Tet-BDP**

In order to analyze the doubly r**TPT3^{CP}**-modified *glmS* ribozyme, the properties of its self-cleaving reaction should be examined. Therefore, **glmS_RNA^{CP}** was incubated with **^{H}Tet-BDP** (**57**) and subsequently submitted to heat annealing in a suitable buffer (section 5.3.14), ensuring a maximum fraction of the RNA to fold correctly. At ambient temperature (~22 °C) cleavage was induced by the addition of glucosamine 6-phosphate (GlcN6P, **1**) to a 5 µM final concentration. The fast cleaving ribozyme was thus kept far below cofactor saturation (ca. 200 µM)[61,67] to enable tracking of the cleavage process within reasonable time frames. Samples were taken at defined time points, directly mixed with a formamide-/ethylenediamine tetraacetic acid (EDTA) solution preventing further cleavage activity and the ribozyme fragments were separated by 12 % DPAGE (Figure 61A). Band intensities were obtained via fluorescence scanning and data of the reaction progress were fitted to a single exponential function.

Cleavage occurred at an average observable rate of k_{obs} = 0.31 ± 0.02 min^{-1} and reached a plateau within 20 min (Figure 61B). Thus, full functionality of the fluorophore-labeled *glmS* ribozyme construct **glmS_RNA^{CP}-Tet-BDP** was confirmed and presented the first-time example of a catalytically active RNA prepared by fusion PCR and *in vitro* transcription with an expanded genetic code. Impressively, introduction of r**TPT3^{CP}** TP (**34**) modifications and subsequent IEDDA click reaction with tetrazine-fluorophore conjugate **57** did not result in decreased activity.

Total cleavage of 70–80 % could be achieved using the r**TPT3^{CP}** labeling approach (see section 5.3.14 for raw data). Along with the reported k_{obs} these values are in good to excellent agreement with literature findings using ^{32}P-endlabeled *glmS* ribozyme constructs[61,62,67,73].

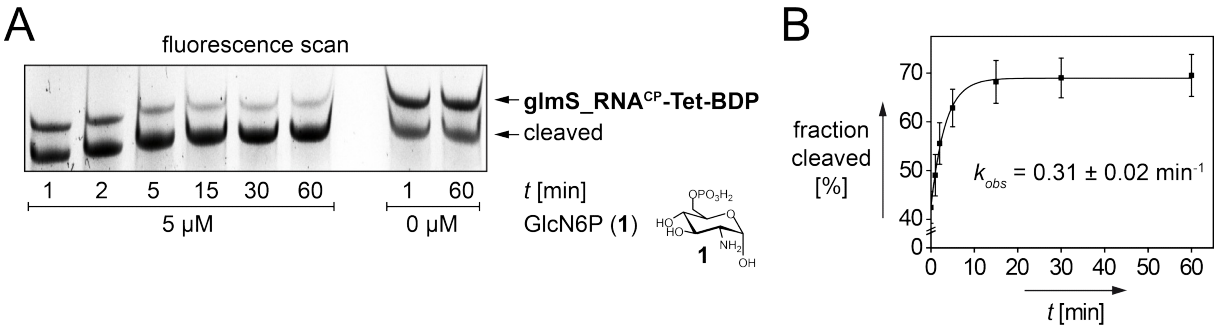


Figure 61. GlcN6P (**1**)-induced self-cleavage of **glmS_RNA^{CP}-Tet-BDP**; A. Representative fluorescence scan of a cleavage assay analyzed via 12 % DPAGE; B. Plot of **glmS_RNA^{CP}-Tet-BDP** self-cleavage, each data point is the average of thirteen individual measurements, error bars are SDs thereof.

Despite *glmS* being the only ribozyme species known so far which needs a cofactor (GlcN6P, **1**) to perform catalysis[70,73] it is notable, that the self-cleavage reaction can also be activated by low millimolar concentrations of tris(hydroxymethyl)aminomethane (tris)[65,67], which are typically contained in PCR and transcription buffers. Therefore, PCR products were desalted prior to transcription and protocols for T7 transcription as well as DNase digestion were adapted to HEPES-based reaction buffers (see 5.3.6 and 5.3.7). Complete exclusion of tris could not be avoided as enzyme stock solutions are usually only available in tris-buffered glycerol.

Interestingly, *in vitro* synthesis of this particular ribozyme construct consistently yielded ~40 % cleaved ribozyme without the addition of any cofactor. This value exceeds the magnitude that would be reasonable for tris-activated self-cleavage. *GlmS* ribozyme cleavage rates and cleavable fractions sensitively depend on the RNA sequence[67], thus an enhanced reactivity of this particular ribozyme construct is conceivable. Further small molecule compounds containing amine functions could also be able to stimulate self-cleavage. Purification of *glmS* transcripts via a preparative DPAGE-electroelution approach (5.3.11) is thus not favorable due to high ammonium acetate concentrations used during the latter process. Gel filtration, membrane filtration or silica-based purification could serve as alternatives. However, crude and DNase-digested transcripts already were of sufficient quality for further biochemical studies such as cleavage assays, which beneficially avoids denaturing conditions and retains the original folding of transcripts[339]. To exclude r**TPT3^{CP}** nucleotides being responsible for premature cleavage of **glmS_RNA^{CP}** and to gather more data characterizing the new RNA constructs, further studies involving HPLC-MS measurements were applied.

HPLC-(ESI$^-$)MS analysis of modularly site-specifically r**TPT3^{CP}**-labeled *glmS* variants

The assembly PCR strategy to generate full-length ribozyme template DNAs could also be adapted to serve as a modular approach. Depending on the application of d**NaM**-modified (**glmS_DNA^{1NaM}**, **glmS_DNA^{3NaM}**) or unmodified antisense segments (**glmS_DNA^{1n}**, **glmS_DNA^{3n}**) in addition to sense strand **glmS_DNA^{2n}**, *glmS* ribozymes could easily be produced with zero (**glmS_RNA^{n}**), one (**glmS_RNA^{CP}1**) or two (**glmS_RNA^{CP}**) site-specific r**TPT3^{CP}** modifications (Figure 62).

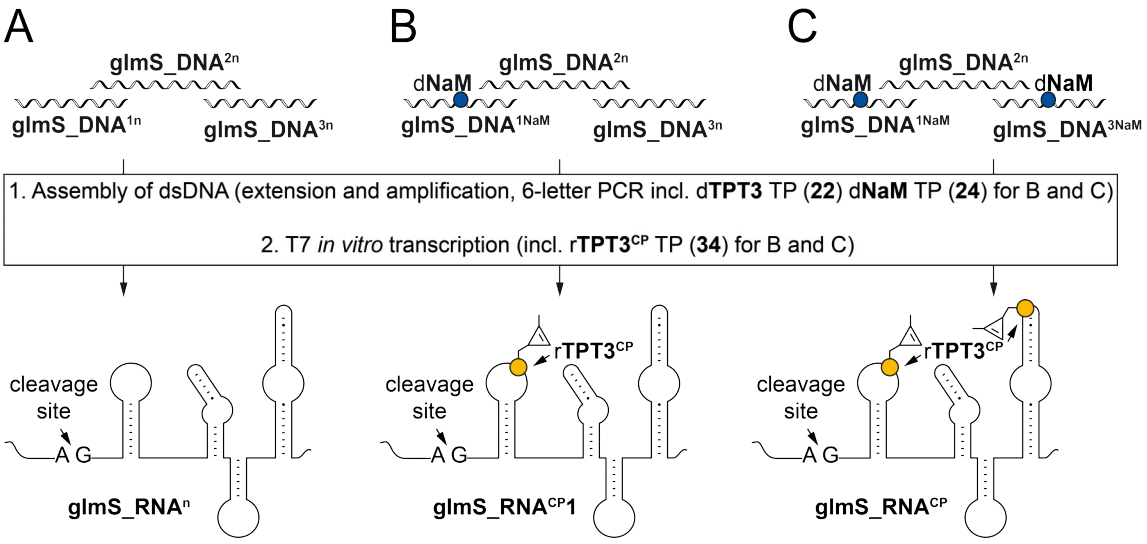


Figure 62. Fusion PCR as modular approach; Construction of unmodified **glmS_RNA^{n}** (A), singly r**TPT3^{CP}**-modified **glmS_RNA^{CP}1** (B) and doubly modified **glmS_RNA^{CP}** (C).

The corresponding three different template DNAs were prepared, transcribed into the respective *glmS* ribozyme variants and analyzed by HPLC-MS. For each ribozyme construct the 5′-cleavage product fragment was detected as 2′,3′-cyclic phosphate (3′-cP) in similar amounts confirming the *glmS* catalytic mechanism (Figure 63, light gray shaded panels). This excludes r**TPT3^{CP}** modifications as suspected origin of premature cleavage. Analyzing the ribozyme core, full and cleaved ribozyme masses were found in good agreement to calculated values, further verifying the presence and identity of desired *glmS* ribozymes (Figure 63, shaded in darker gray).

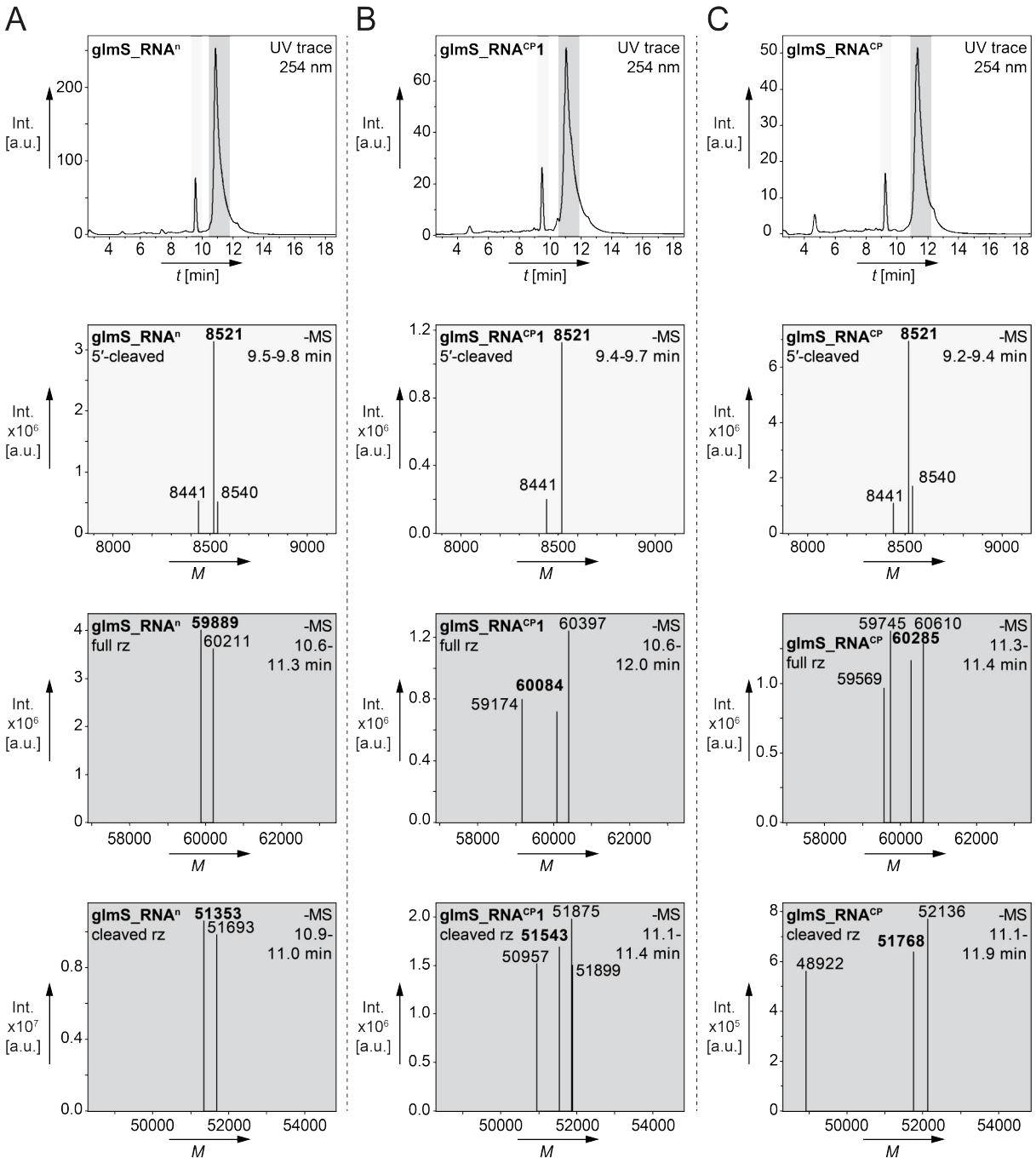


Figure 63. HPLC-(ESI$^-$)MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) of modularly generated *glmS* ribozyme variants showing (top down) UV trace and deconvoluted spectra of the 5′ cleaved fragment ($M_{calcd.}$ = 8522 for [M 5′-TP-3′-cP+Na$^+$]), full ribozyme (rz), and cleaved rz for **glmS_RNA^n** (A, $M_{calcd.}$ = 59898 for [M full rz], $M_{calcd.}$ = 51353 for [M cleaved rz]), **glmS_RNA^{CP}1** (B, $M_{calcd.}$ = 60093 for [M full rz], $M_{calcd.}$ = 51548 for [M cleaved rz]) and **glmS_RNA^{CP}** (C, $M_{calcd.}$ = 60289 for [M full rz], $M_{calcd.}$ = 51767 for [M cleaved rz]); most accurate assignments are highlighted in bold.

Analysis of large *in vitro* transcribed RNA molecules such as the 185 nt *glmS* constructs via electron spray ionization mass spectrometry (ESI-MS) is not widely-used and straight-forward evaluation of the nucleic acid data is hindered by several factors. The infeasibility to accurately separate full ribozyme from its cleaved core sequence by HPLC, formation of

multi sodium ion adducts even after desalination of the samples, and the ability of T7 polymerase to produce transcripts with significant heterogeneity of 5′- and 3′-ends[48] is challenging. HPLC-ESI MS is a powerful tool to efficiently and precisely determine molecular weights of oligonucleotides of ≤150 nt[39,349–351], yet literature is focusing on synthetic nucleic acids, which indicates that analysis of *in vitro* prepared RNAs is far more demanding. Fortunately, accurate mass spectra could be obtained employing a defined and identical set of deconvolution parameters to all produced ribozyme variants (see section 5.3.13 for details). Apart from single events, assignment of expected species was generally possible within a ±5 *m*/*z* range, which represents a satisfying span for molecules of this size. Complete correlation of the values appearing in the deconvoluted mass spectra of Figure 63 to RNA molecules can be found in Table 1.

Table 1. Assignment of calculated ($M_{calcd.}$) and experimentally evident (M_{found}) masses from ESI$^-$ mass spectrometric analyses of *glmS* variants **glmS_RNA^n**, **glmS_RNA^{CP}1**, and **glmS_RNA^{CP}**; most accurate masses are highlighted in bold.

***glmS* fragment**	$M_{calcd.}$	M_{found}	identification
cleaved fragment (3′-cP)	**8442**	**8441**	**5′-DP + Na$^+$**
	8522	**8521**	**5′-TP + Na$^+$**
	8545	**8540**	**5′-TP + 2Na$^+$**
glmS_RNA^n	**59898**	**59889**	**full rz; 5′-TP + 2Na$^+$**
	60204	60211	full rz + A; 5′-TP + Na$^+$
	51353	**51352**	**cleaved rz; 5′-OH**
	51682	51693	cleaved rz + A; 5′-OH
glmS_RNA^{CP}1	59176	59174	full rz - 3′-3 nt; 5′-TP + 2Na$^+$
	60093	**60084**	**full rz; 5′-TP + 2Na$^+$**
	60400	60397	full rz + A; 5′-TP
	50959	50957	cleaved rz - 3′-3 nt; 5′-OH
	51548	**51543**	**cleaved rz; 5′-OH**
	51877	51875	cleaved rz + A; 5′-OH
	51900	51899	cleaved rz + A; 5′-OH + Na$^+$
glmS_RNA^{CP}	59574	59569	full rz -3′-2 nt; 5′-DP
	59746	59745	full rz - 3′-2 nt; 5′-TP + 4Na$^+$
	60289	**60285**	**full rz; 5′-TP + Na$^+$**
	60618	60610	full rz + A; 5′-TP + Na$^+$
	48928	48922	cleaved rz - 3′-9 nt+ Na$^+$
	51767	**51768**	**cleaved rz; 5′-OH**
	52142	52136	cleaved rz + A; 5′-OH + 2Na$^+$

Strikingly, no masses of unmodified **glmS_RNA^{n}** or singly r**TPT3^{CP}**-modified **glmS_RNA^{CP}1** were detected for the *glmS* ribozyme with two modifications (**glmS_RNA^{CP}**). This strongly supports the observation that incorporation of r**TPT3^{CP}** nucleotides via T7 *in vitro* transcription is efficient, specific, and can be used for site-specific labeling of long RNA fragments.

Analyzing the modular *glmS* constructs via electrophoretic mobility shift assay (EMSA)

Investigating and purifying biomolecules, biotinylation and subsequent streptavidin binding is a popular and highly efficient method.[160,166,352] A tetrazine-biotin conjugate (**58**, Figure 64A) was synthesized by C. Domnick in the Kath-Schorr group for applications on strained alkene-modified RNAs.[327,337] Ribozyme constructs **glmS_RNA^{n}**, **glmS_RNA^{CP}1**, and **glmS_RNA^{CP}** modularly bearing either 0, 1, or 2 r**TPT3^{CP}** modification/s, respectively, were submitted to a biotinylation and streptavidin-binding assay. Therefore, the crude and DNase-digested transcripts were incubated with tetrazine **58** for 1 h at room temperature followed by gel filtration. Subsequent streptavidin treatment was performed for 1 h at ambient temperature and the samples were analyzed on DPAGE in comparison with unreacted RNA (Figure 64B, detailed experimental settings can be found in section 5.3.9). As expected, unmodified **glmS_RNA^{n}** does not give any response to this approach. For **glmS_RNA^{CP}1** and **glmS_RNA^{CP}** UV bands shifted towards lower electrophoretic mobility are observed. These bands correspond to biotinylated and streptavidin-bound species, showing the applicability of modularly r**TPT3^{CP}**-modified RNAs in biotin-streptavidin EMSA approaches. **glmS_RNA^{CP}1** exhibits two lower migrating bands, likely belonging to a full and cleaved ribozyme species in biotinylated and streptavidin-bound status. For **glmS_RNA^{CP}** four bands can be detected, which suggests singly and doubly clicked RNA each of the full and cleaved ribozyme construct. Unreacted species are still visible on DPAGE analysis. No attempts were made to optimize any parameters of the click reaction, streptavidin binding, or evaluation. Crude transcripts have been involved in this qualitative experiment, definitive RNA concentrations were not determined. Therefore, it is possible that tetrazine **58** was applied in a sub-stoichiometric amount. A large excess tetrazine-biotin conjugate (and subsequent removal of abundant reagent by gel filtration) could potentially facilitate full conversion. Future attempts could further employ agarose gel electrophoresis since the properties of DPAGE might partially break the strong biotin-streptavidin interaction thus inhibiting exact quantification.

The presented results demonstrate successful modular approach to site-specifically label RNA with r**TPT3^{CP}** functions susceptible of IEDDA click reactions and might facilitate detecting and capturing target RNAs out of a complex context such as cell lysates.

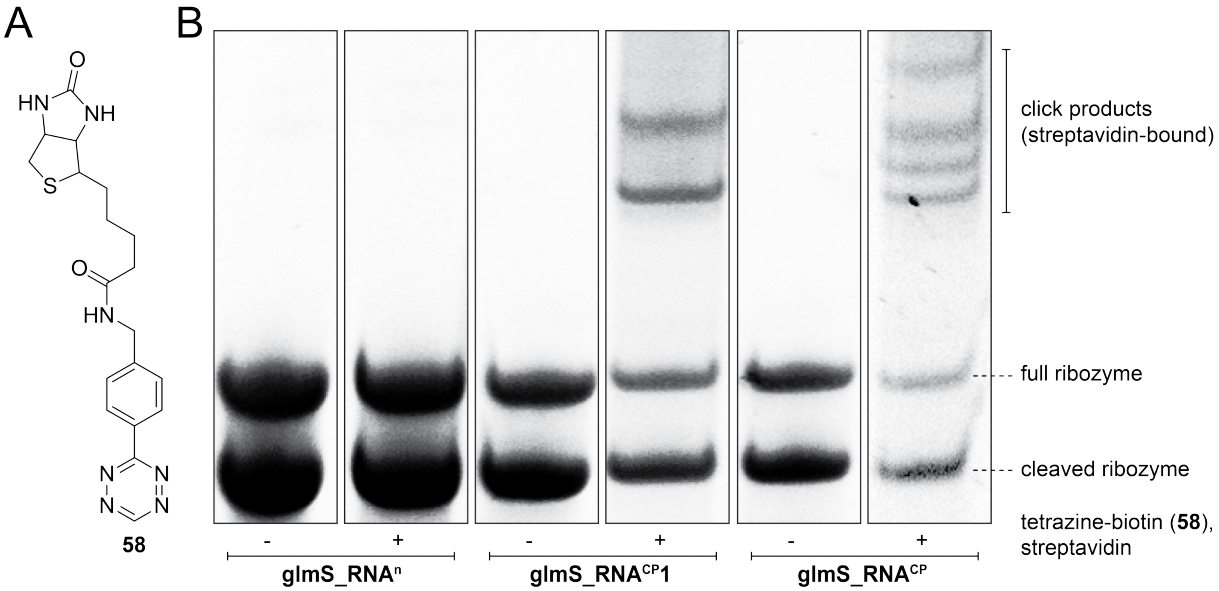


Figure 64. Applying an EMSA assay on r**TPT3^{CP}**-modified *glmS* constructs; A. Chemical structure of tetrazine-biotin conjugate (**58**, synthesized by C. Domnick[235]); B. 8 % DPAGE analysis (*SYBR Safe* stain) of crude *glmS* transcripts with no (**glmS_RNA^{n}**), one (**glmS_RNA^{CP}1**), or two (**glmS_RNA^{CP}**) r**TPT3^{CP}** labels, each before and after incubation with tetrazine **58** and subsequent streptavidin treatment; full gel shown in section 7.4, Figure 331.

Using a well-studied *glmS* sequence as model system, it could be shown that site-specific introduction of one or two unnatural r**TPT3^{CP}** nucleotides by *in vitro* transcription from fusion PCR-generated dsDNA templates containing unnatural d**NaM**:d**TPT3** base pair(s) enables preparation of modified and still fully functional ribozymes. These examples of catalytically active RNAs produced via genetic alphabet expansion transcription show the high potential of this approach for studying complex RNA molecules.

The fusion PCR strategy using three overlapping single stranded (ssDNA) strands impressively expands site-specific r**TPT3^{CP}** labeling to longer RNAs. Assuming a limit of 180 nt for solid-phase DNA template generation (*biomers.net, sigmaaldrich.com,* accessed 2019/07) and taking overlap regions of ca. 30 nt as well as loss of a ~20 nt T7 promoter region into account, site-specifically labeled ribonucleic acid molecules of about 450 nt in length should be easily accessible by this approach.

The presented technique generally allows labeling of long non-coding RNAs (lncRNAs) which are arbitrarily defined to be of more than 200 nt in length, many of them easily comprise >1 kb[60,116,353,354]. In order to establish site-specific r**TPT3^{CP}** labeling for such lncRNAs exceeding the currently available sequence length, new strategies had to be developed. The results aiming on this goal will be shown in the following section.

3.1.2.3 Labeling of long RNAs (in cooperation with P. Leifels[354])

Extension of the labeling approach via augmented genetic alphabet transcription was intended to facilitate site-specific r**TPT3^{CP}** RNA labeling of virtually any length.

As exemplary sequence an r**TPT3^{CP}** nucleotide should be incorporated into the A region part of the 17 kb *Xist* lncRNA[75–78,83,84] using genetic alphabet expansion transcription and visualized by subsequent IEDDA cycloaddition with a tetrazine-fluorophore conjugate.

In this case, generation of the d**NaM**-containing template DNA should be established starting from isolated plasmid DNA. Therefore, P. Leifels investigated in his bachelor thesis the implication of a forward primer consisting of a sequence complementary to a particular region of a *Mus musculus Xist* gene-inserted plasmid DNA as well as an overhang containing the T7 promoter information. Creating linear dsDNA from the plasmid, several reverse primers were examined regarding their suitability for efficient PCR amplification. After successful application of unmodified reverse primers a set of d**NaM**-modified analogs causing initial d**NaM**:dA mismatches when hybridized with plasmid DNA were tested to introduce unnatural bases necessary for downstream r**TPT3^{CP}** labeling of the *Xist* RNA A region.

Further optimization efforts resulted in the identification of the primer pair **Xist_Primer_FW** and **Xist_Primer_RV^{NaM}** fulfilling the required efficient replication as well as introduction of T7 RNA polymerase promoter and a d**NaM**:d**TPT3** unnatural base pair during PCR (Figure 65). *In vitro* transcription from the modified dsDNA template in the presence of r**TPT3^{CP}** TP (**34**) was envisioined to prepare a cyclopropene-decorated 401 nt *Xist* RNA fragment (**Xist_RNA^{CP}**). Subsequent IEDDA reaction with a tetrazine-fluorophore conjugate (**^{Me}Tet-BDP**, **59**) was planned to give site-specifically fluorophore-tagged **Xist_RNA^{CP}-Tet-BDP**.

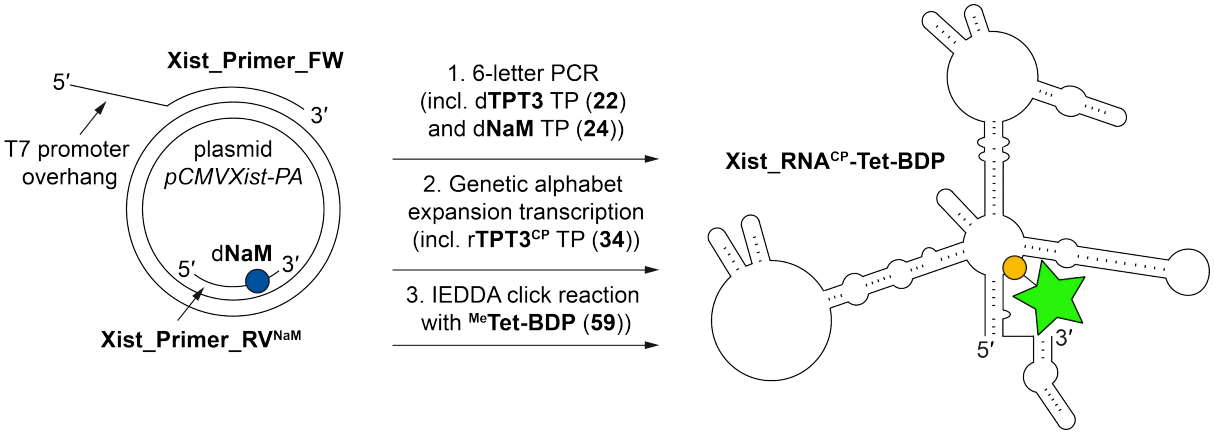


Figure 65. Schematic representation of site-specific r**TPT3^{CP}** labeling of a *Xist* RNA fragment via *in vitro* transcription of a plasmid-derived DNA template and subsequent IEDDA click reaction with a tetrazine-fluorophore conjugate.

During the course of this work final experiments on the *Xist* project were conducted. The DNA template for *in vitro* transcription was prepared by PCR amplifying the *Xist* A region, in particular nucleotides 365 to 766 of plasmid *pCMVXist-PA* (see section 5.3.5 for details) using **Xist_Primer_FW** and **Xist_Primer_RV^{NaM}** allowing for a d**A**:d**NaM** mismatch (plasmid position 738) in the plasmid:primer hybrid.. A 401 nt long *Xist* A region fragment was subsequently generated via *in vitro* transcription in the presence of r**TPT3^{CP}** TP (**34**) giving **Xist_RNA^{CP}**. Control transcript **Xist^{n}** was prepared in parallel using unmodified reverse primer **Xist_Primer_RV^{n}**. IEDDA reaction with 6-methyl substituted **^{Me}Tet-BDP** (**59**) at the cyclopropene moiety at position 738 resulted in the fluorescently labeled *Xist* fragment of **Xist_RNA^{CP}** (Figure 66A). Two fluorescent bands could be detected by DPAGE of the crude, unpurified *Xist* transcript post-transcriptionally decorated with a fluorophore. One migrated at the expected height of ~400 nt and a second band exhibited much slower migration behavior. **Xist_RNA^{n}** consisting of canonical nucleotides did not give a fluorescent response when incubated with **59** but showed the same pattern upon DPAGE separation (Figure 66B).
Native 1 % agarose gel electrophoresis of **Xist^{CP}** (Figure 66C) yielded only one visible band at the expected height. This finding supports that the slow migrating species on the denaturing polyacrylamide gel is caused by a highly stable superstructure of the r**TPT3^{CP}**-modified *Xist* RNA (**Xist_RNA^{CP}**) that is incompletely denatured.

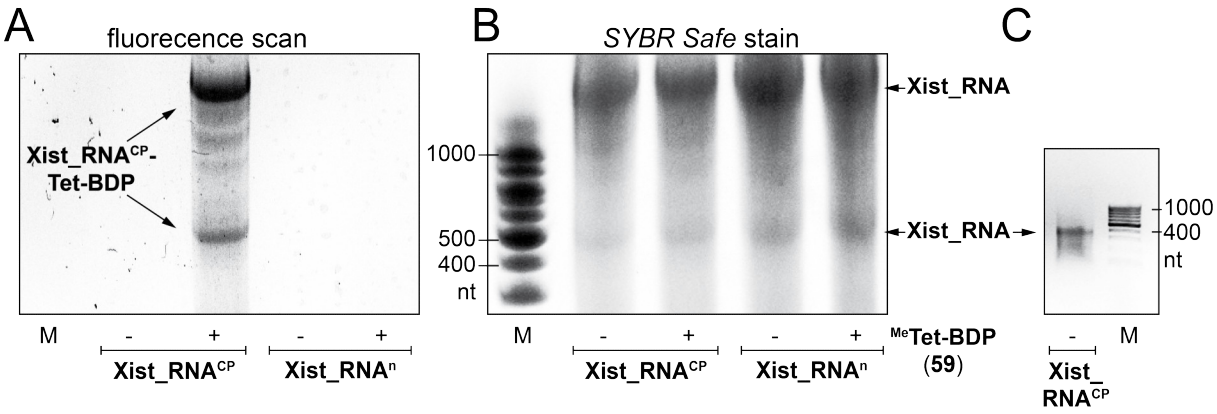


Figure 66. Gel electrophoretic analysis of *Xist* A region RNA constructs; A. Fluorescence scan of a 10 % DPAGE analysis with fluorescent bands of **Xist_RNA^{CP}-Tet-BDP**; B. *SYBR Safe* post staining of the same gel; C. 1 % agarose electrophoresis (EtBr stain) of **Xist_RNA^{CP}**.

Extensive structure formation of certain RNA sequences is often not completely unfolded although strong denaturing conditions are applied before and during gel electrophoresis. Migration behavior on polyacrylamide gels can be strongly affected resulting in electrophoretic mobility anomalies that are not observed applying agarose gel electrophoresis.[355]

Notably, efficient fluorescent labeling of this lncRNA could be achieved without the need for prior purification of the transcript. Laborious and potentially denaturing purification steps[356] could be avoided and serve as an advantage studying large functional non-coding RNAs.
This approach represents a pioneering step enabling site-specific r**TPT3^{CP}** labeling of long and highly structured target RNAs for post-transcriptional attachment of reporter groups using additive-free and bio-orthogonal IEDDA cycloaddition reactions, which can even be performed in live cells[234,357–359]. The presented techniques ensure an ultra-mild fluorophore introduction, thus it is ideally suited to fluorescently label virtually any long RNA molecule with complex folding[70,71] and allows to maintain the original RNA structure which arises from *in vitro* transcription. Most lncRNAs are at least several hundred nucleotides in length and site-specific labeling via *in vitro* transcription represents an advantageous approach to prepare fluorescence-tagged large RNAs for biochemical and biophysical studies.
As the number of actual molecule copies decreases when handling *in vitro* transcribed long RNA samples, sensitive fluorescent readout might be limited at some point. Generating RNA constructs of increased length, incorporation of multiple UBP modifications in the plasmid DNA template should be considered. This demands carefully considered construct design since d**NaM**:dN mismatch repeats will probably not be tolerated to an infinite number by polymerases and will decrease melting temperatures.
This setup provides a facile way to introduce an unnatural base pair into any sequence derived from plasmid DNA at both 5′- and 3′-ends. More sophisticated techniques involving scission and recombination of desired plasmid DNAs[360–362] should potentially also allow facile internal introduction of unnatural base pairs into any sequence of choice.

3.1.2.4 Incorporation efficiency and specificity of 34 into RNA

When introducing unnatural nucleotides such as the r**TPT3^{CP}** TP (**34**) into RNA standard T7 RNA polymerase was employed (see section 5.3.6 for amino acid sequence). The enzyme was not genetically engineered to process UB triphosphates and reaction conditions were not optimized in favor of UB TP incorporation; therefore, questions on conversion efficiency and fidelity arose.
The Romesberg group reported superior efficiency and specificity for *in vitro* transcription from **DNA^{NaM}** employing the ribonucleoside triphosphate r**5SICS** TP.[187] They further found that the performance of the UBP **TPT3**:**NaM** was generally outranging the **5SICS**:**NaM** base pair in PCR replication.[190,193,194,202,363] Extrapolating these data to RNA level, outstanding incorporation efficiency during transcription is expected. Nonetheless, various experiments were executed to analyze the enzymatic recognition properties of the functionalized r**TPT3^{CP}** nucleotide **34**.

Relative amount of RNA produced incorporating r**TPT3^{CP}** nucleotides

Overall RNA production was compared employing templates **DNA^{n1}** to incorporate canonical nucleotides and **DNA^{NaM}** as template to introduce one r**TPT3^{CP}** nucleotide in the presence of **34** and rNTPs (Table 2). Impressively, the transcription from d**NaM**-modified DNA results in an averaged ~1:4.5 ratio compared to the native RNA, stating that T7 RNA polymerase is able to efficiently produce significant amounts of r**TPT3^{CP}**-modified oligonucleotides. These numbers clearly outrange the literature-known turnover when r**5SICS** nucleotides are incorporated into RNA using d**NaM**-containing template DNAs. The Romesberg group reported a 20-fold reduction in transcription efficiency for this particular system[187]. Therefore, the **TPT3** nucleobase analog was found to represent a significantly advanced alternative to **5SICS** and performance is not affected by the cyclopropene linker.

Table 2. Comparison of RNA yields obtained from transcription reactions employing only canonical rNTPs or rNTPs and r**TPT3^{CP}** TP (**34**); crude reactions were purified by preparative DPAGE and electroelution prior to concentration determination; values are the average of at least two independent trials (single values can be found in Table 24, section 5.3.12).

transcript	n [nmol]	normalized
RNA^{n1}	4.1 ± 1.2	100 ± 29 %
RNA^{CP}	0.9 ± 0.2	22 ± 4 %

Transcription yields from **DNA^{NaM}** were further evaluated comparing the incorporation of r**TPT3^{CP}** TP (**34**) with previously developed r**TPT3^{NOR}** TP[235] (**29**) (Table 3). Data of 30 min and 5 h reaction time were evaluated and resulted in a similar trend. As expected, the bulkier norbornene-modified r**TPT3^{NOR}** nucleotide is incorporated with lower efficiency than the r**TPT3^{CP}** derivative bearing the less sterically demanding cyclopropene linker. While also exhibiting faster reaction kinetics, r**TPT3^{CP}** incorporation into RNA represents a delightful improvement to currently available labeling techniques employing an expanded genetic alphabet.

Table 3. Comparing incorporation efficiency of r**TPT3^{CP}** TP (**34**) and r**TPT3^{NOR}** TP (**29**) during T7 *in vitro* transcription of **DNA^{NaM}**; each data point is the average of three independent trials, errors are standard deviations (SDs) thereof; see section 5.3.12 for single values.

	r**TPT3^{CP} TP (34)**		r**TPT3^{NOR} TP (29)**	
	30 min	5 h	30 min	5 h
n [nmol]	0.17 ± 0.05	0.64 ± 0.18	0.09 ± 0.03	0.44 ± 0.25
Normalized	100 %	100 %	53 %	69 %

Relative amounts of full-length **RNA^{CP}** produced from transcription of **DNA^{NaM}**

Even though not being observed by unspecific RNA staining employing DPAGE analysis, a closer look on HPLC-MS measurements of crude **RNA^{CP}** transcripts revealed minor presence of a truncated RNA species (Figure 67, peak at ~6.8 min). The molecular weight corresponds to a transcript stopping at the unnatural base site. Integrating the UV traces of five independent measurements, the amount of full-length **RNA^{CP}** transcript was found to be 87±3 % (see 5.3.13 for single values).

This is still an extraordinary high rate and does not affect the potential of r**TPT3^{CP}** TP (**34**) to be a very promising candidate for site-specific modification of RNA during *in vitro* transcription from d**NaM**-containing DNA templates, even with a linker system attached to the nucleobase analog scaffold.

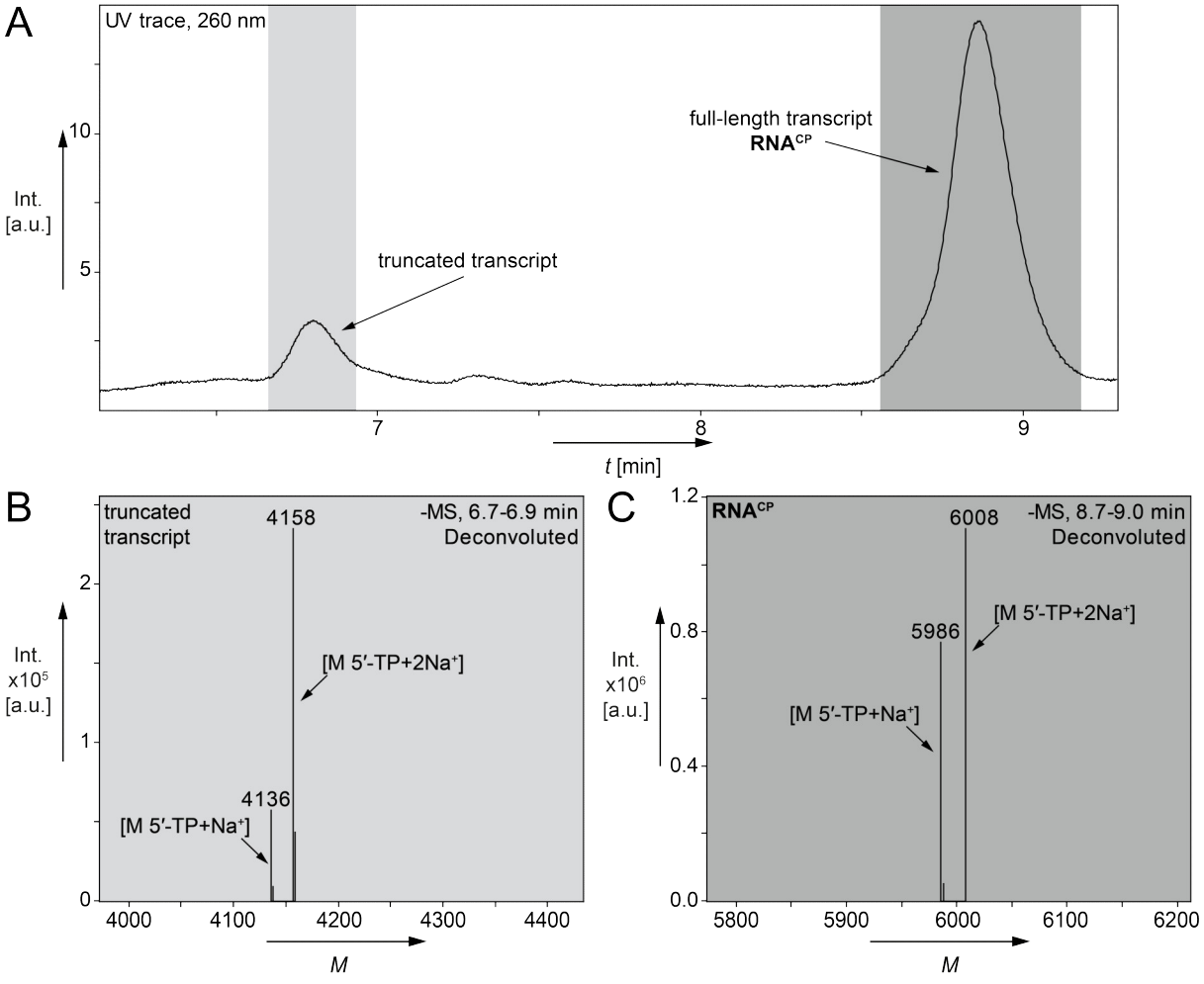


Figure 67. Close-up view of HPLC-MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) of a crude **RNA^{CP}** transcription; A. UV trace; B. Deconvoluted ESI$^-$ mass spectrum of the truncated transcript peak eluting at ~6.8 min ($M_{calcd.}$ = 4135 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 4157 for [M 5′-TP+2Na$^+$]) C. Deconvoluted ESI$^-$ mass spectrum of the full-length **RNA^{CP}** ($M_{calcd.}$ = 5985 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 6007 for [M 5′-TP+2Na$^+$]).

For short transcripts such as the presented **RNA^{CP}**, truncated and full-length transcripts are separable either via DPAGE or HPLC-MS; therefore, access to in-depth information about efficiency and obtained nucleic acid species is given. This does, however, not account for longer RNAs as limiting resolution will restrain from clearly quantifying a ratio of full-length to shorter RNAs.

As mentioned above, the short transcript length of **RNA^{CP}** (17 nt) is in close approximation to the initiation phase of T7 RNA polymerase (≤12 nt)[48] and the unnatural r**TPT3^{CP}** nucleotide is incorporated at position 13. During this stage, the enzyme is known to naturally fall off the template frequently.[364,365] Facing an unnatural nucleotide which is presumably more demanding, an increased truncation event is thus comprehendible. In posterior elongation phase of the transcription, procession is not inhibited as easily. Transcriptions yielding **tRNA^{CP}** or r**TPT3^{CP}**-modified *glmS* variants did not result in detection of a truncated RNA during DPAGE analysis, strongly supporting this hypothesis.

Considerable variability in transcription efficiency was observed using r**TPT3^{CP}** TP (**34**) of different batches. Therefore, most possible caution is recommended when performing synthesis, purification and aliquotation because highest quality of the unnatural cognate ribose triphosphate **34** is crucial for excellent labeling success.

Specificity of expanded genetic alphabet transcription from d**NaM**-modified DNA

To gain further insights into enzymatic recognition and procession of the UBP **TPT3**:**NaM** by T7 RNA polymerase, a transcription reaction with d**NaM**-modified template **DNA^{NaM}** was run in the absence of cognate r**TPT3^{CP}** TP (**34**) (Figure 68). When the enzyme reaches the unnatural base site in the template, three possibilities are envisioned. The polymerase could fall off the template DNA lacking a suitable nucleotide to incorporate, leaving a truncated RNA. Skipping the d**NaM** site would result in a transcript shortened by one nucleotide. Further, mismatching the UB nucleotide with a canonical nucleotide would give a full-length transcript. Also a mixture of all scenarios is imaginable.

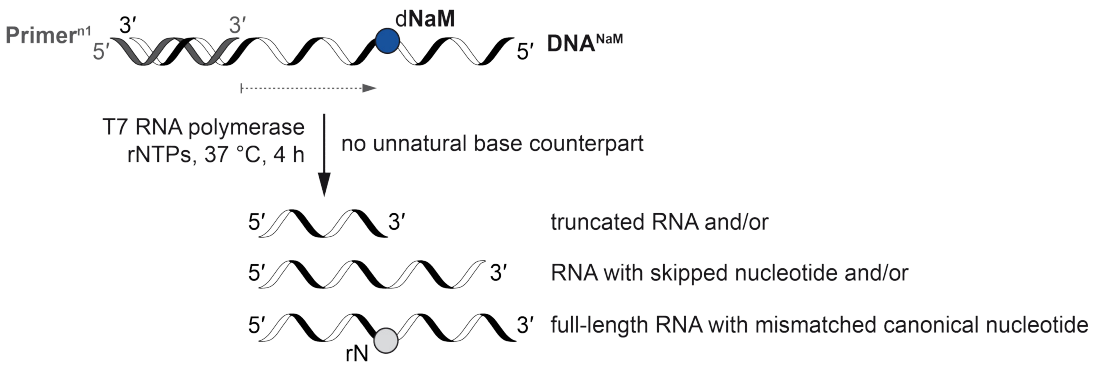


Figure 68. Schematic representation of a T7 *in vitro* transcription from d**NaM**-containing template **DNA^{NaM}** with reaction outcome for stalling, skipping, or mismatch introduction events when encountering the UB site.

A transcription reaction was prepared under standard conditions (see 5.3.6), desalted via gel filtration and subjected to HPLC-MS analysis (Figure 69). Two major peaks are observed in a ~1:1 ratio. The species eluting at ca. 7.5 min corresponds to a truncated fragment before the d**NaM** site in the template. This confirms the assumption that T7 RNA polymerase is able to discriminate the UBP from native base pairing and specifically needs a cognate nucleotide for transcription.

The second peak, eluting at ca. 8.3 min, in contrast could be assigned to a full-length RNA transcript containing either an r**A** or r**U** nucleotide, stating that the enzyme is partially able to mispair the **NaM** nucleobase with a canonical building block. Unfortunately, an unambiguous attribution was not feasible within the accuracy of the employed method. The mass of an adenine nucleobase anion ($M_{calcd.}$ = 134.045 for $[M]^-$) interestingly closely fits the mass of uridine in adduction with a sodium ion ($M_{calcd.}$ = 134.009 for $[M+Na^+]$). RNA transcripts are commonly detected as multiple sodium ion adducts using ESI, even after proper desalination; therefore, high resolution mass spectrometric methods would be needed to distinguish between the two species. Misincorporation of both r**A** and r**U** canonical nucleotides opposite d**NaM** cannot be excluded at this point. Notably masses of transcripts corresponding to an r**G** or r**C** nucleotide in the sequence were not observed.

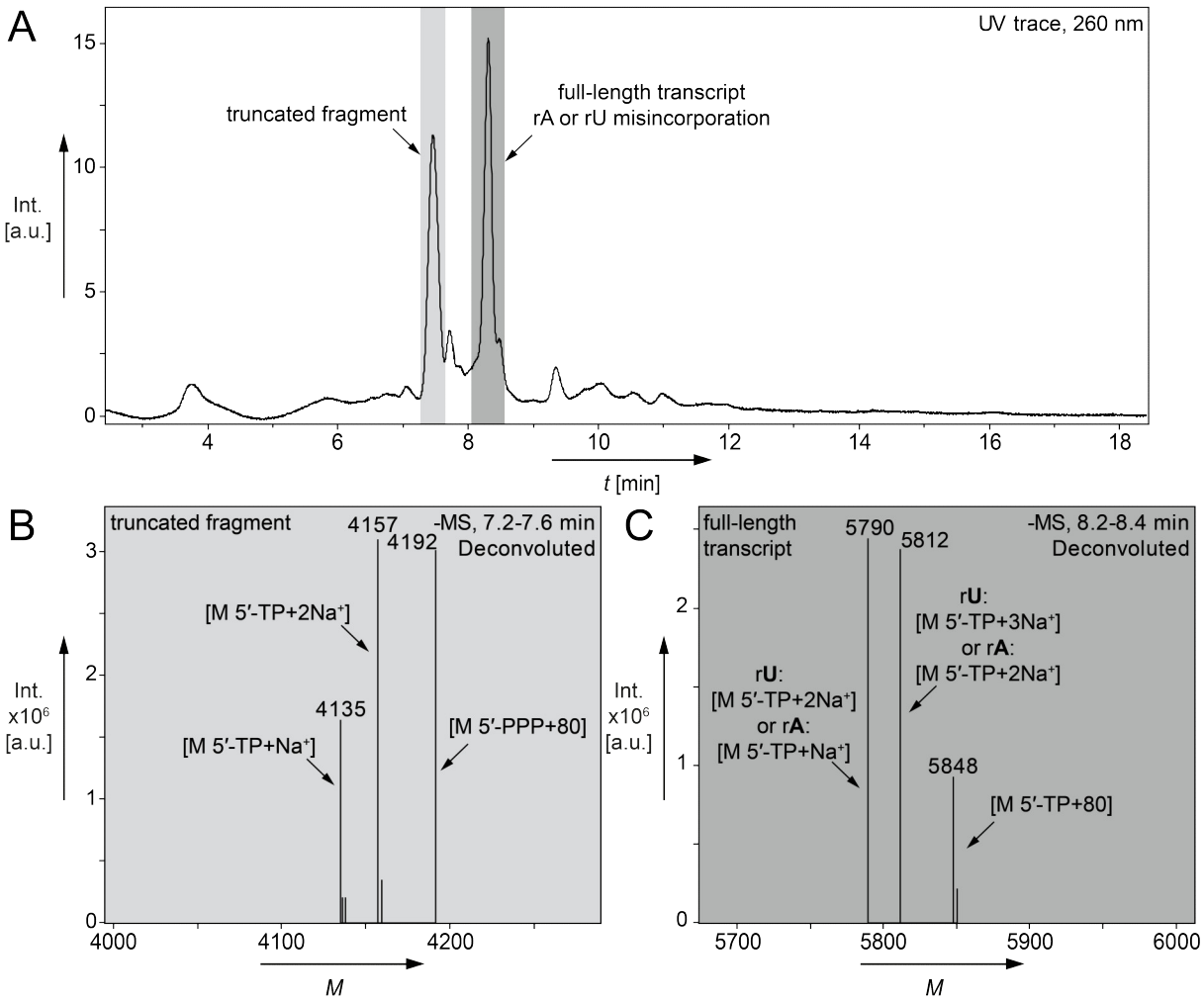


Figure 69. HPLC-MS analysis (3 % MeCN for 9 min followed by a gradient 3→40 % MeCN/buffer in 11 min, 0.2 mL min^{-1}, column *Xterra*, see 5.3.13) of a transcription from d**NaM**-modified DNA template **DNA^{NaM}** in the absence r**TPT3^{CP}** TP (**34**); A. UV trace; B. Deconvoluted ESI$^-$ spectrum of the truncated transcript eluting at ~7.4 min ($M_{calcd.}$ = 4135 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 4157 for [M 5′-TP+2Na$^+$]); C. Deconvoluted ESI$^-$ spectrum of full-length transcripts containing either misincorporated r**U** or r**A** (r**U**: $M_{calcd.}$ = 5789 for [M 5′-TP+2Na$^+$], $M_{calcd.}$ = 5811 for [M 5′-TP+3Na$^+$]; r**A**: $M_{calcd.}$ = 5790 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 5812 for [M 5′-TP+2Na$^+$].

The presented experiments clearly lead to the conclusion, that r**TPT3^{CP}** TP (**34**) is incorporated into RNA via *in vitro* transcription with outstanding efficiency and specificity. T7 RNA polymerase readily accepts the unnatural nucleotide **34** as cognate substrate opposite d**NaM** nucleotides in template DNA and produces high amounts of modified RNA with low error rate. Only in the absence of r**TPT3^{CP}** TP (**34**) the enzyme is partly able to misincorporate canonical nucleotides into the elongating ribonucleic acid.

3.1.2.5 Efficiency of IEDDA reactions on rTPT3^{CP}-modified RNA

It was previously shown by the Kath-Schorr group, that *in vitro* IEDDA cycloadditions on strained alkene-modified RNAs are possible with near-quantitative yield by just applying a slight excess (≤5 equivalents) of tetrazine reagents.[234,235] In former experiments, short chemically synthesized or r**TPT3^{NOR}**-labeled sequences were investigated. High reactivity of r**TPT3^{CP}** nucleotides incorporated into RNA is expected and should be analyzed.

Using the short 17 nt sequence **RNA^{CP}** a band shift towards lower mobility could be observed upon click reaction with a tetrazine-fluorophore conjugate, allowing to estimate the efficiency of the click reaction. In Figure 54 (see 3.1.2.1) **RNA^{CP}** was incubated with tetrazine-Oregon Green-conjugate **28** in a 1:1 ratio for 1 h and traces of unreacted RNA were still visible. Secondary and tertiary structure formation might partially hinder the fast IEDDA kinetics. Overnight reaction with 7 eq. **28** led to approx. 75 % conversion to **RNA^{CP}-Tet-OG** in an overnight reaction as could be seen from the HPLC-MS analysis shown in section 3.1.2.1, Figure 53.

High conversion rates of click reactions are preferred for biochemical applications such as in-cell experiments. Interestingly, IEDDA click reaction of **RNA^{CP}** with tetrazine **28** (5 eq.) for 1 h at ambient temperatures did not result in any visible starting material applying DPAGE (Figure 70A, left panel). It is, however, arguable that the dark shades resulting from **28** overlap with leftover unreacted RNA, preventing a reliable conclusion. In a comparative experiment employing BODIPY-based **^{H}Tet-BDP** (**57**, 3 eq.) the reaction outcome clearly states full conversion within only one hour (Figure 70A, right panel). Looking at the chemical structures of the fluorophores bound to the tetrazine core (Figure 70B) Oregon Green is considerably more rigid and bulkier than the BODIPY residue which could result in steric hindrance of the IEDDA reaction. Therefore, an enhanced reactivity of **^{H}Tet-BDP** (**57**) towards large and structured biomolecules can be assumed.

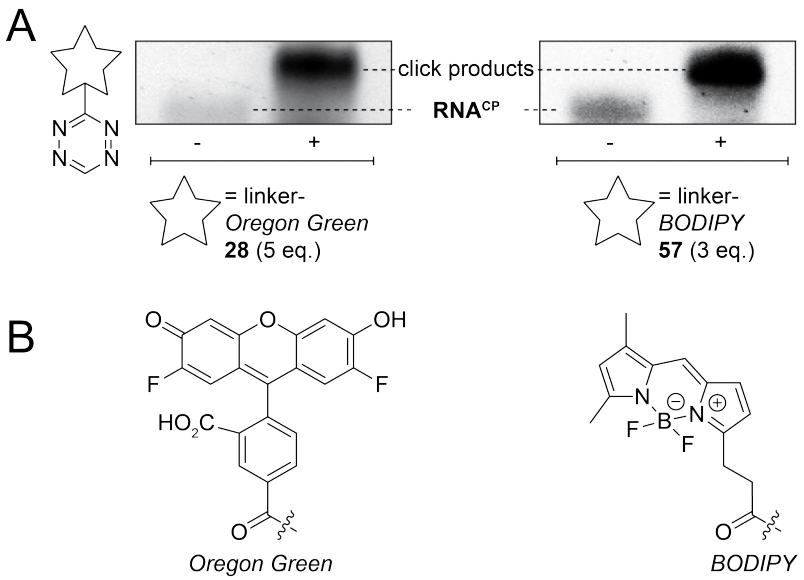


Figure 70. IEDDA click reactions of purified **RNA^{CP}** with tetrazine-Oregon Green conjugate **28** and BODIPY-related tetrazine **57**; A. 20 % DPAGE analyses (*SYBR Safe* staining) showing quantitative conversion (r.t., 1 h); B. Chemical structures of the fluorophore moieties.

With these data it could be shown that quantitative IEDDA click reactions on **RNA^{CP}** are very well feasible even using just a slight excess of reagent. The choice of tetrazine conjugate might have significant impact on reaction dynamics. Quantification of the click reaction is easily shown for small RNAs by gel electrophoretic shift or by liquid chromatographic separation. This does, however, not account for larger sequences. Therefore, a large (at least 10-fold) excess of click reagent is recommended when full conversion of IEDDA cycloaddition is an essential prerequisite.

3.1.2.6 Non-templated incorporation of 34 into short RNAs

A major concern working with unnatural base pairs is the general specificity of incorporation, *i.e.* the question whether UB nucleotides would be inserted into RNA by T7 transcription without a cognate partner in the template. Such events could result in 5′-, 3′- or internal unspecific labeling. While non-templated introduction of unnatural nucleotides within the sequence is not expected as the **TPT3**:**NaM** base pair was carefully optimized (on DNA level) to widely exclude mispairing events with canonical nucleotides[195], unspecific addition of nucleotides at positions beyond templates, *i.e.* the 5′- and 3′-termini, might be a considerable factor. During the course of this study, several events gave reason to investigate unspecific introduction of r**TPT3^{CP}** nucleotides into RNA. The following section addresses this complex issue on various levels.

Assessing non-templated incorporation via HPLC-MS analysis

DNA^{n1} was transcribed in the presence of r**TPT3^{CP}** TP (**34**) similarly to the control experiment mentioned before (section 3.1.2.1), desalted via gel filtration, and subjected to HPLC-MS analysis. An RNA species with unspecifically incorporated r**TPT3^{CP}** nucleotides would potentially result in a shifted peak of the oligonucleotide towards higher retention times. Both the crude ESI$^-$ trace as well as deconvoluted spectra would clearly distinguish between the expected **RNA^{n1}** sequence and other transcription products. Figure 71 shows the UV chromatogram (upper panel) revealing a broadened peak. Integration of the corresponding ESI$^-$ data over the complete course of this signal (lower left panel) gave a distinct oligonucleotide pattern that exclusively yielded the molecular weight of the unmodified **RNA^{n1}** transcript upon deconvolution (lower right panel). These data strongly suggest that non-templated r**TPT3^{CP}** incorporation does not occur using **DNA^{n1}** as template.

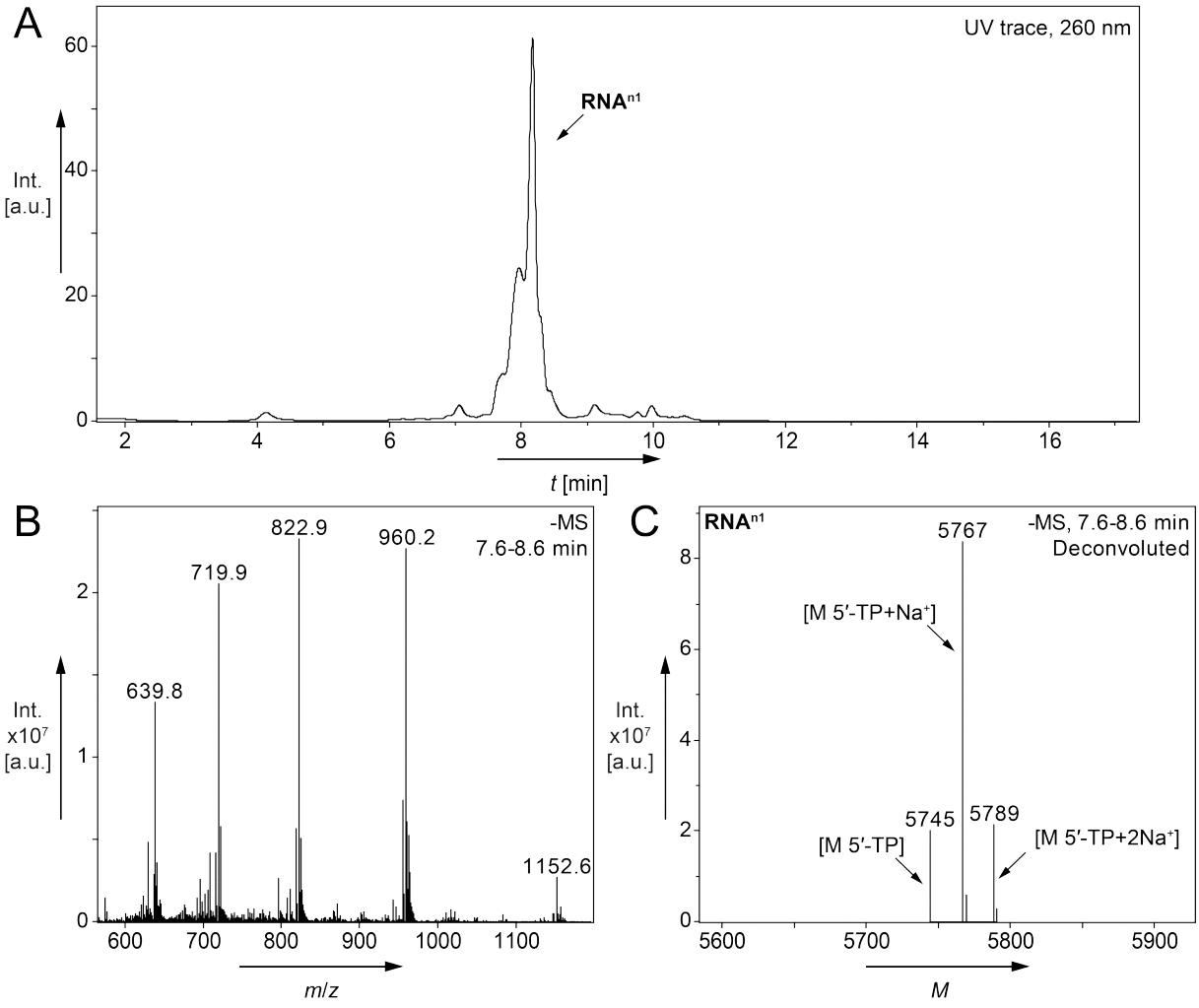


Figure 71. HPLC-MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) of a transcription reaction from **DNA^{n1}** containing only canonical bases in the presence of r**TPT3^{CP}** TP (**34**); A. UV trace; B. Crude ESI$^-$ signals integrated over the whole peak width; C. Deconvoluted ESI$^-$ spectrum of the peak eluting at 7.6–8.6 min ($M_{calcd.}$ = 5745 for [M 5′-TP], $M_{calcd.}$ = 5767 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 5789 for [M 5′-TP+2Na$^+$]).

Applying fluorescent readout on DPAGE to address non-templated r**TPT3^{CP}** nucleotide incorporation

Further experiments were performed to rule out unspecific introduction of unnatural nucleotides into RNA by T7 *in vitro* transcription. As r**TPT3^{CP}** TP (**34**) nucleotides incorporated into RNA allows for efficient post-transcriptional decoration with fluorophores via IEDDA cycloaddition with suitable tetrazines, facile detection of such nucleic acid species should be feasible by fluorescent readout. Therefore, **RNA^{c}** and **tRNA^{c}** were prepared by transcription from unmodified templates **DNA^{n1}** or **tDNA^{n}**, respectively, in the presence of **34** under standard conditions. Resulting ribonucleic acids were incubated with an excess (4-fold) tetrazine-Oregon Green conjugate **28** and analyzed using DPAGE via fluorescence scan and comparative *SYBR Safe* stain (Figure 72).

Both **RNA^{c}** and **tRNA^{c}** gave intense bands when submitted to unspecific nucleic acid stain (Figure 72, lower panels). Fluorescence scanning the gel reveals no significant signal in either case. Arguably, a hardly detectable faint band could be noticed, yet fluorescent readout has a very low detection limit thus the intensities could not be clearly related. An unspecific non-covalent binding of tetrazine **28** to RNA is conceivable in a concentration-dependent manner. Unspecific incorporation of r**TPT3^{CP}** TP (**34**) is considered occurring, if at all, to an insignificant extent for these constructs.

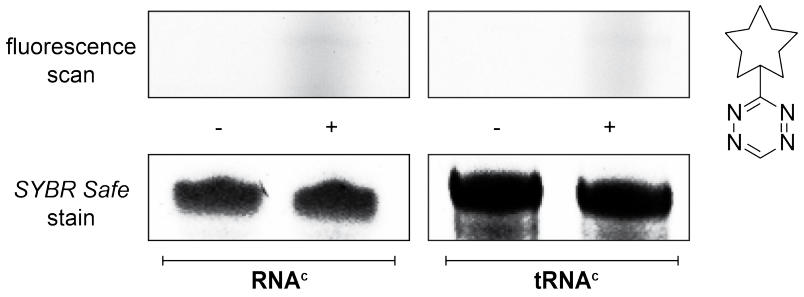


Figure 72. 20 % DPAGE analysis of control transcription reactions using the unmodified, natural-base DNA templates in the presence of r**TPT3^{CP}** TP (**34**) for *in vitro* transcription; resulting **RNA^{c}** and **tRNA^{c}** were incubated with 4 eq. tetrazine-fluorophore conjugate **28**; fluorescence scan of the gel (upper panels) and *SYBR Safe* post-staining (lower panels).

Determining unspecific **34** incorporation opposite canonical bases using a fluorogenic assay

In addition, an even more sensitive approach was developed to address non-templated enzymatic procession of unnatural triphosphate **34** by T7 RNA polymerase. Unmodified **DNA^{n1}** and d**NaM**-modified **DNA^{NaM}** were transcribed with and without the addition of r**TPT3^{CP}** TP (**34**).

The obtained RNA was subjected to a fluorescent assay (Figure 73). First, the crude, DNase-digested samples were purified by gel filtration to remove excess **34** which would have false positive impact on the downstream readout process. **RNA^{CP}** and **RNA^{n1}** were subsequently incubated with a 10-fold excess of tetrazine-fluorophore **57** facilitating complete IEDDA cycloaddition to proceed. Further gel filtration was performed to minimize background signal of **^{H}Tet-BDP** (**57**). Samples were analyzed using the fluorescence mode of a plate reader.

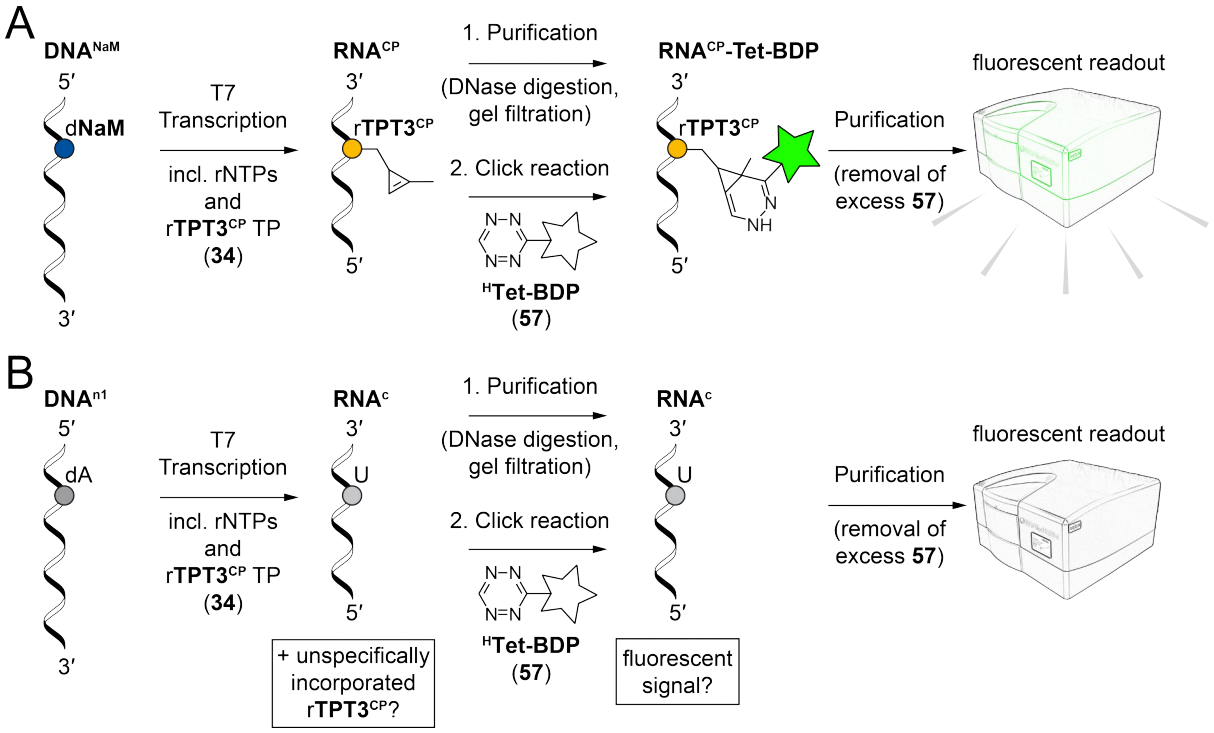


Figure 73. Schematic representation of a plate reader-based fluorescent assay assessing unspecific r**TPT3^{CP}** incorporation during *in vitro* transcription; A. Modified **DNA^{NaM}** results in **RNA^{CP}**, subsequent IEDDA reaction with **57** gives turn-on fluorescence of **RNA^{CP}-Tet-BDP** which is detectable via fluorescent readout of the samples; B. **DNA^{n1}**, submitted to the same procedure, does not increase fluorescence.

Analyzing the transcript of unmodified template **DNA^{n1}** no turn-on fluorescence could be detected with increasing r**TPT3^{CP}** TP (**34**) concentrations from 0.1 to 2.0 mM added to the transcription reaction (Figure 74). For the positive control reaction with **DNA^{NaM}**, where efficient incorporation of the cyclopropene-modified unnatural nucleotide has been demonstrated, the fluorescence intensity was strongly increased (approx. 100-fold) upon incubation with **^{H}Tet-BDP** (**57**). Misincorporation of r**TPT3^{CP}** opposite natural nucleotides was not observed within the detection limits of this fluorescence-based assay and remained at a negligible range for this sequence.

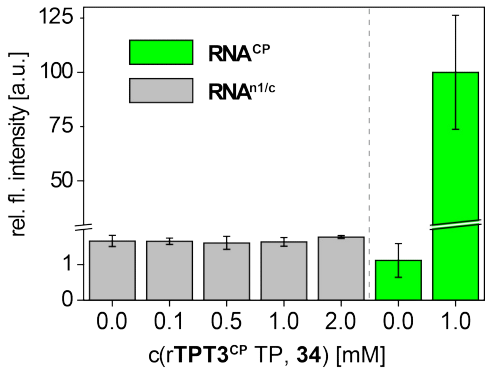


Figure 74. Evaluation of a fluorometric assay determining unspecific r**TPT3^{CP}** TP (**34**) incorporation during T7 *in vitro* transcription from **DNA^{n1}** in comparison with **DNA^{NaM}**; histogram of relative fluorescence intensities detected for **RNA^{n1}**/**RNA^{c}** (grey bars) and **RNA^{CP}** (green bars) in the presence of varying concentrations of **34** and subsequent incubation with **^{H}Tet-BDP** (**57**); data are the average of at least two independent measurements, background-subtracted and normalized (see section 5.3.15).

Together with the mass spectrometric analysis as shown above (section 3.1.2.1, Figure 52), which resulted only in detection of the expected molecular mass values for a singly r**TPT3^{CP}**-modified RNA, these sound investigations clearly exclude significant non-templated incorporation opposite natural nucleotides or at the 5′/3′-end during T7 *in vitro* transcription from **DNA^{NaM}**. These observations might still be depending on each particular sequence context, sequence length, or the actual unnatural nucleotide employed for transcription as well as applied incubation times, concentration of ribose triphosphates, and potentially involved reporter systems such as tetrazine-fluorophore conjugates.

3.1.2.7 Non-templated 3′-extension of 34 into longer RNAs

K. Kulikov employed site-specific r**TPT3^{CP}** labeling on the *CPEB3* ribozyme (89 nt), proving versatility of the approach to produce functional RNAs (section 3.1.2.1).[343,344] With an r**TPT3^{CP}** modification near the 5′-end of the ribozyme, the cleaved ribozyme core unintentionally appeared faintly fluorescing after incubation with **^{Me}Tet-BDP** (**59**) in some occasions. This was also true for RNA transcribed from the unmodified template in the presence of r**TPT3^{CP}** TP (**34**). The ribozyme's 5′-cleaved fragment did not show any fluorescence under these conditions, which led to the conclusion that unspecific incorporation took place in the 3′-region of the ribozyme.

Furthermore, P. Leifels encountered unspecific incorporation of r**TPT3^{CP}** nucleotides to a perceivable amount using fluorescence readout of DPAGE analysis during his bachelor thesis on a 401 nt fragment of the *Xist* A region.[354]

T7 RNA polymerase is known to extend transcript molecules by one or two nucleotides both at the 5′- and 3′-end in a non-templated manner.[48,366] Depending on the sequence, 3′-extension can affect greater than 50 % of all transcribed RNA copies as described in the literature.[48] Therefore, significant unspecific addition of r**TPT3^{CP}** nucleotides is conceivable.
The collection of presented data suggests that non-templated elongation of transcripts is an issue to be taken seriously applying expanded genetic alphabet *in vitro* transcription with r**TPT3^{CP}** triphosphate **34**. Though this phenomenon could be excluded for reactions with template **DNA^{n1}** and **tDNA^{n}** (section 3.1.2.6), it is assumed to be a common feature using T7 RNA polymerase with particular relevance for the production of longer constructs. This might be explained by the difference in polymerase behavior during initiation and elongation stage.
Strategies to avoid unspecific 3′-incorporation of unnatural nucleotides could involve the addition of small self-cleaving ribozymes to the target RNA sequence (Figure 75). For example, the well-known HDV ribozyme (~50 nt)[106,367] could be encoded in the template DNA downstream the RNA of interest. The catalytically active RNA would be generated during *in vitro* transcription and readily undergo Mg^{2+}-dependent scission of a specific upstream phosphodiester bond homogenously producing a 3′-cP. Unspecifically incorporated nucleotides would be appended to the HDV sequence and could thus be removed during purification of the target RNA. This approach, however, decreases the length of accessible target sequences from synthetic DNA. Additionally it has been found, that RNAs of interest can interfere with formation of essential structural elements of the ribozyme thus lowering cleavage activity.[48] To apply ribozyme-mediated production of defined 3′-ends careful design of template DNA as well as target RNA is necessary and can potentially suffer from drawbacks.

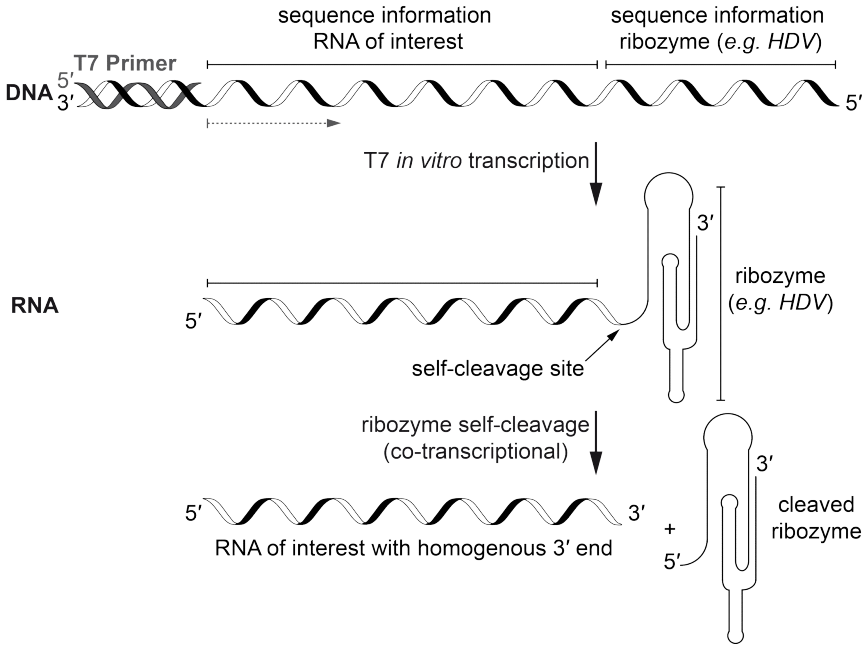


Figure 75. Generation of RNA with homogenous 3′-terminus via *in vitro* transcription using a self-cleaving ribozyme downstream the sequence of interest (reproduced from Beckert and Benoît[106]).

Another very promising method to suppress non-templated incorporation of nucleotides to the 3′-terminus of RNA during T7 *in vitro* transcription is the application of 2′-methoxy(OMe)-modifications at the 5′-terminus of antisense template strands.[368,369] Approaching the end of template sequence, binding stability of T7 RNA polymerase is significantly decreased, forcing the enzyme to fall off and end RNA synthesis. This technique is easily accessible for transcription from synthetic DNA and can also be applied for PCR-based template preparation employing 2′-OMe-modified primer sequences. Utilization of such is recommended for investigations on expanded genetic alphabet transcription if 3′-homogenous RNA is essential to avoid ambiguous data. Further investigations on this topic are presented as part of the discussion in section 3.2.2.1.

3.1.2.8 Site-specificity of rTPT3^{CP} *in vitro* incorporation

Ensuring site-specific incorporation of unnatural entities such as r**TPT3^{CP}** nucleotides is a prerequisite for downstream biochemical applications of transcribed RNA.

The d**NaM** position in short DNA templates is ensured by chemical solid-phase synthesis of the oligonucleotides. Preparation of dsDNA templates modified with d**TPT3**:d**NaM** UBPs via PCR relies on efforts and developments by the Romesberg group, who extensively demonstrated outstandingly high efficiency and fidelity during replication by standard DNA polymerases.[193,194,332,363]

Earlier studies by C. Domnick provided evidence of site-specific r**TPT3^{NOR}**-UB incorporation.[327,337] Reverse transcription from a modified RNA resulted in stalling at the

unnatural nucleotide site. Site-specific introduction of r**TPT3**$^{\mathbf{NOR}}$ nucleotides was thus clearly verified. Analogous properties of r**TPT3**$^{\mathbf{CP}}$ can be concluded, since the unnatural nucleobases differ only in the linker system, which is not part of the base pair recognition.

A cleavage-deficient *CPEB3* ribozyme variant bearing an r**TPT3**$^{\mathbf{CP}}$ nucleotide at the scissile site was prepared by K. Kulikov[343], demonstrating high accuracy of template-directed r**TPT3**$^{\mathbf{CP}}$ TP procession (**34**).

This thesis additionally excluded misincorporation of r**TPT3**$^{\mathbf{CP}}$ opposite natural nucleobases in template DNAs during *in vitro* transcription by several reliable control experiments on short RNAs in (see sections 3.1.2.1, 3.1.2.4, and 3.1.2.6).

HPLC-MS experiments on modularly assembled *glmS* ribozymes bearing zero, one, or two r**TPT3**$^{\mathbf{CP}}$ nucleotides revealed exclusive molecular masses for the respective RNA constructs, guaranteeing incorporation events only to occur at templated sites even for longer sequences (section 3.1.2.2). An EMSA assay could additionally distinguish between the singly and doubly labeled *glmS* variants, which also supports site-specificity.

Furthermore, a powerful approach based on reverse transcription to determine the position and to estimate abundance of unnatural bases in RNA was developed in cooperation with K. Kurscheidt and E. Hoffmann.[370–372] Section 3.4.3 will present and discuss this technique in detail.

The extensive results presented and discussed throughout the past sections have shown in-depth preparation of the novel unnatural nucleotide r**TPT3**$^{\mathbf{CP}}$ TP (**34**) and detailed analysis of applications in T7 *in vitro* transcription from d**NaM**-modified DNA templates.

r**TPT3**$^{\mathbf{CP}}$ nucleotides can be introduced into RNA via T7 *in vitro* transcription with high efficiency and exclusive accuracy in considerable quantities. Broad novel application possibilities encouraged for further implementations with r**TPT3**-derived nucleotides, which will be presented in the following sections.

3.2 PELDOR distance measurements in RNA via site-specific spin labeling using r**TPT3** derivatives

For a thorough understanding of functional non-coding RNA molecules, knowledge about their three-dimensional conformation is essential. Structural investigations on highly structured RNA molecules, however, are challenging and require sophisticated techniques. For example, crystallization of transcribed RNAs is often not easily achieved due to heterogeneity, insufficient purity, or intrinsic flexibility and thus remains rate-limiting in the analysis of large RNAs.[48,87,90,91] Additionally the obtained data represent single conformations at potentially high salt concentrations which might not accurately reflect the solution state. Chemical probing may be limited with respect to steric hindrance or even impact target structure due to hybridization events with the molecule of interest.[109,117,118,120,373]

In contrast, spin labeling offers a powerful alternative to elucidate dynamics and structural motifs of biomolecules under physiological conditions.[259,263,374] Averaged spectra obtained from RNAs bearing nitroxide radical spin labels enable major contributions studying nucleic acids by the means of electron paramagnetic resonance (EPR) spectroscopy involving inter-nitroxide distances by pulsed electron-electron double resonance (PELDOR) experiments.[123,127,342]

In the following, collaboration projects with G. Hagelueken and O. Schiemann will be presented. Recording of all EPR measurement data as well as *MtssIWizard* modeling as performed and provided by G. Hagelueken.

3.2.1 Post-transcriptional RNA spin labeling by IEDDA reactions

Efficient and site-specific introduction of r**TPT3^{CP}** nucleotides into RNA via *in vitro* transcribed RNA was extensively shown during this study. C. Domnick set out to adapt this promising technique for post-transcriptional attachment of nitroxide groups to RNA molecules.[327,375] During his studies, a suitable tetrazine derivative bearing a 2,2,5,5-tetramethyl-3-carboxylpyrrolin-1-oxyl (TEMPYO) spin label, **Tet^{NO}** (**60**, Figure 76A) was synthesized. By means of the general template-directed T7 transcription protocol (section 3.1.2.1)[337,344,376] employing d**NaM**-containing template **DNA^{sc_NaM}**, a self-complementary r**TPT3^{CP}**-modified sequence of 18 nt was prepared, **RNA^{sc_CP}** (Figure 76B), which would self-hybridize giving an RNA duplex. The **RNA^{sc_CP}** dimer was post-transcriptionally conjugated with spin labels via IEDDA click reactions using **Tet^{NO}** (**60**).

Continuous wave (*cw*)-EPR spectroscopic analysis of **RNA^{sc_CP}-Tet^{NO}** (Figure 76C) revealed reasonable immobilization of the nitroxide and spin counting gave an excellent labeling

efficiency (94 %). This once again impressively proved efficient incorporation of r**TPT3^{CP}** nucleotides into RNA by template-directed *in vitro* transcription in combination with an outstandingly high conversion applying IEDDA cycloadditions on r**TPT3^{CP}**-modified RNA oligonucleotides.

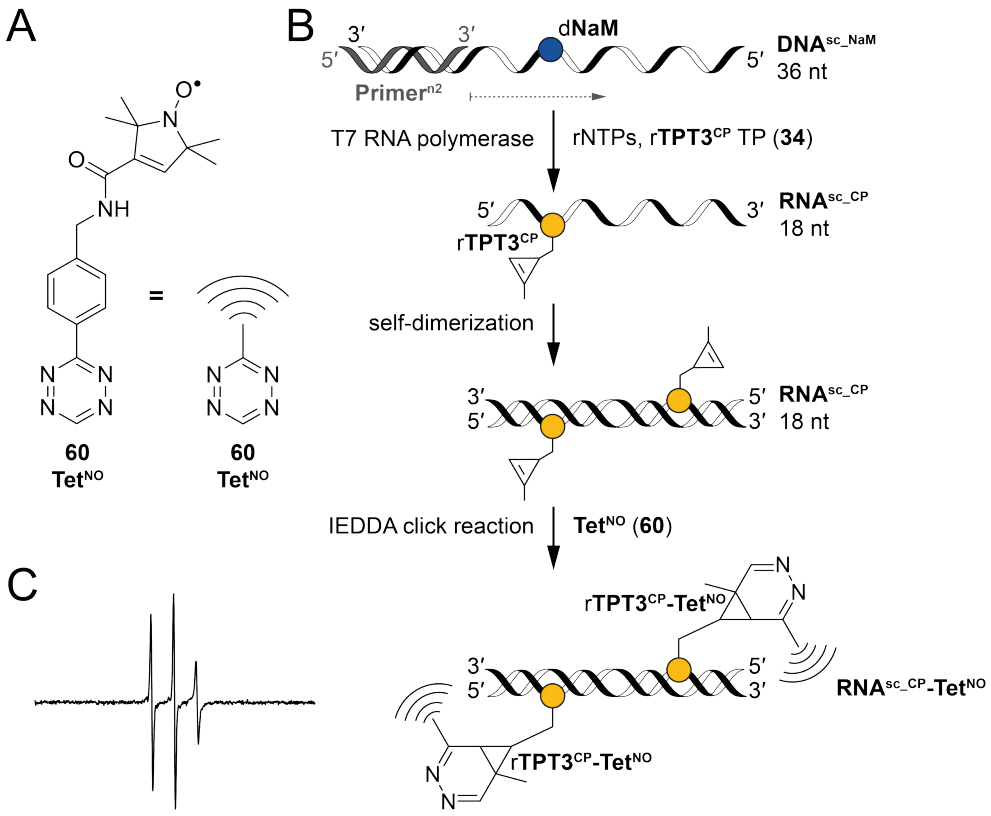


Figure 76. Post-transcriptional spin labeling of a self-complementary RNA; A. Chemical structure of **Tet^{NO}** (**60**); B. Schematic representation of the spin labeling approach using site-specific r**TPT3^{CP}** nucleotide introduction and subsequent IEDDA click reaction with **60**; C. *cw*-EPR spectrum of clicked **RNA^{sc_CP}-Tet^{NO}**; adapted from Domnick *et al.*[327,375]

PELDOR distance measurements on **RNA^{sc_CP}-Tet^{NO}** yielded a well-shaped probability density curve with a maximum at 6.1 nm (Figure 77A, black curve). Molecular dynamics (MD) simulations by S. Kath-Schorr were in excellent agreement to the obtained distance distribution (Figure 77A, shaded in green), confirming the experimental data. The major peak appeared somewhat broadened compared to other spin labeling techniques based *e.g.* on solid-phase synthesis[264] or post-synthetic application of click chemistry on alkyne-modified oligonucleotides[127]. These approaches benefit from rather rigidly attached spin labels. The IEDDA cycloaddition product used by C. Domnick *et al.* resulted in a considerably long and flexible linker system with many degrees of freedom. Advantageously, this method can potentially serve to produce very large, *in vitro* prepared RNA constructs which are not easily addressable by other techniques.

During the course of this thesis, further validation of **RNA^{sc_CP}-Tet^{NO}** was performed. Surprisingly, UV melting experiments of **RNA^{sc_CP}-Tet^{NO}** (Figure 77B) did not result in altered T_m of the modified duplex in comparison to unmodified **RNA^{sc_n}** (68.4 ± 0.5 °C vs. 67.1 ± 0.3 °C, respectively). This was not expected since the r**TPT3^{CP}** entities cause a total of two mismatches in the duplex and usually lowering of the melting temperature is observed disturbing the Watson-Crick base pairing. An MD structure snapshot suggested a certain ability of the cyclopropene-pyridazine-TEMPYO linker system to fold back on the RNA covering the ends of the duplex, potentially re-stabilizing the dimer. Since the self-complementary sequence also allows for intramolecular hairpin formation, this equilibrium could sensitively influence experiments such as T_m determination. Concentration and nature of salt additives and oligonucleotides as well as annealing procedures are major factors determining the ratio of duplex to hairpin structure.[377–379] Thus, gathering reliable data from self-complementary RNA can be more challenging than anticipated.

Analysis via DPAGE (Figure 77C) exhibited a monomer as well as a duplex band for both **RNA^{sc_n}** and **RNA^{sc_CP}-Tet^{NO}** even under strong denaturing conditions, which implies very stable duplex formation, supporting the findings in melting temperature experiments.

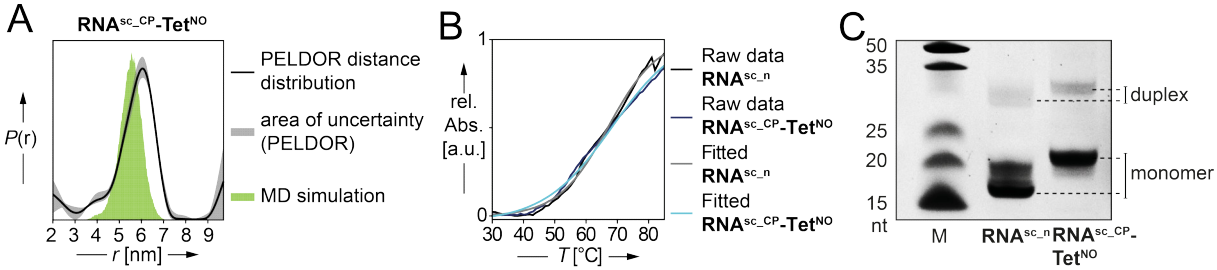


Figure 77. Analysis of TEMPYO spin-labeled duplex **RNA^{sc_CP}-Tet^{NO}**; A. PELDOR distance distribution (adapted from Domnick *et al.*)[375] with area of uncertainty (gray) caused by background correction, and MD-simulated probability (by S. Kath-Schorr, green); B. UV melting curves of unmodified duplex **RNA^{sc_n}** (T_m = 67.1 ± 0.3 °C) and spin-labeled dimer **RNA^{sc_CP}-Tet^{NO}** (T_m = 68.4 ± 0.5 °C); C. 20 % DPAGE analysis (*SYBR Safe* stain) of the unmodified RNA and r**TPT3^{CP}**-modified duplex RNA clicked with tetrazine **60**.

Conclusively, the method presented by C. Domnick *et al.* serves as a useful technique to study *e.g.* global RNA-RNA or RNA-protein interactions with potential applicability for in-cell measurements[337]. However, aiming on more precise experimental data, *i.e.* aspiring a high-resolution nanometer-range ruler for RNA investigations, further advances had to be taken which will be discussed in the next section.

3.2.2 Site-specific RNA spin labeling employing the novel nitroxide-derivatized r**TPT3^{NO}** TP (**61**)

Approaching the need to introduce a more rigid spin label linker system to be site-specifically introduced into RNA, C. Domnick synthesized and optimized a nitroxide-bearing r**TPT3** nucleotide.[327] The novel compound enables the incorporation of spin labels to a sequence of interest directly via *in vitro* transcription. As a compromise between rigidity and T7 incorporation efficiency using d**NaM**-containing DNA templates, he developed r**TPT3^{NO}** TP (**61**, Figure 78).

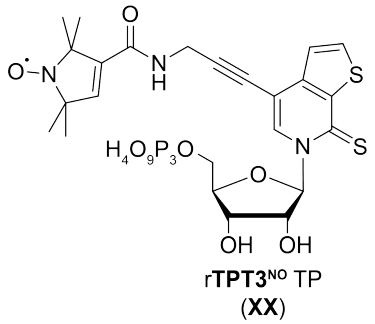


Figure 78. Chemical structure of novel nitroxide-functionalized r**TPT3^{NO}** TP (**61**) synthesized by C. Domnick[327].

3.2.2.1 Labeling of short self-complementary sequences

C. Domnick performed first experiments implementing site-specific incorporation of r**TPT3^{NO}** nucleotides into RNA.[327] The following section implements both these initial results and various achievements of continued efforts as part of this work.

18-mer duplex **RNA^{sc_NO}**

Submitting d**NaM**-containing template **DNA^{sc_NaM}** (see section 3.2.1) to a T7 transcription reaction in the presence of r**TPT3^{NO}** TP (**61**) C. Domnick was able to generate self-complementary **RNA^{sc_NO}** by *in vitro* transcription according to established protocols[235,376] (Figure 79).

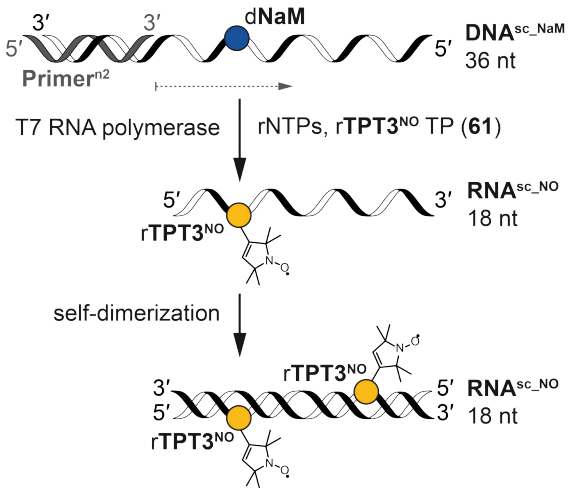


Figure 79. Schematic approach to *in vitro* prepare self-complementary **RNA^{sc_NO}** duplex containing two TEMPYO spin labels, reproduced from C. Domnick[327].

A fairly immobilized *cw*-EPR spectrum (Figure 80A, reproduced during this thesis work) demonstrated successful application of the novel r**TPT3** derived triphosphate **61** in site-specific spin labeling of RNA. Showing considerable modulation (Figure 80B) the corresponding PELDOR experiment yielded a distance distribution of the **RNA^{sc_NO}** self-dimer (Figure 80C, black line). The defined spectrum exhibited a sharp distance distribution with a maximum at 4.97 nm. *MtsslWizard*[380] calculations by G. Hagelueken placed the unnatural base in a Watson-Crick geometry with an opposite r**A** mismatch and were in extensive agreement with the measurement (Figure 80C, shaded in gray; model shown on the right). The distance is in close approximation to a similar RNA duplex published by Kerzhner *et al*.[127] who used copper-catalyzed azide-alkyne click reactions to attach spin labels post-synthetically to RNA, confirming versatility of the presented transcription-based technique.

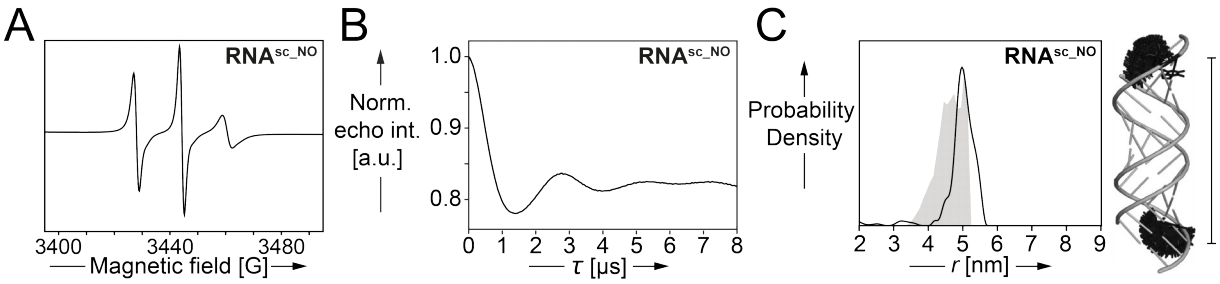


Figure 80. EPR spectroscopic analysis of **RNA^{sc_NO}**; A. *cw*-EPR spectrum; B. Background-corrected fit of the echo time-trace (for raw data see section 5.3.18); C. PELDOR distance distribution (black line) in comparison to model calculations (gray area, model shown on the right); Data of B/C obtained by C. Domnick[327]);.

20-mer duplex **RNA^{sc_ext_NO}**

With these promising first results, this work should further evaluate resolution and reliability of the new RNA spin labeling approach. An extended template DNA (**DNA^{sc_ext_NO}**, see 5.3.19), synthesized by *Ella Biotech* (Martinsried, Germany), resembled the same sequence as in **DNA^{sc_NO}** but comprised additional two building blocks (dG and dC) in the middle of the oligonucleotide. In the resulting (again self-complementary) extended transcript **RNA^{sc_ext_NO}** the nitroxide labels of incorporated r**TPT3^{NO}** nucleotides were thus placed in a distance elongated by two bases that would add ~0.5 nm to their proximity.

The sequence was transcribed in a similar manner as shown for the short self-dimerizing **RNA^{sc_NO}** (see Figure 79). *cw*-EPR spectroscopy revealed favorable immobilization of spin labels (Figure 81A). Rather moderate modulation depth could be observed in the corresponding PELDOR echo time trace (Figure 81B), which might be explained by increased hairpin structure formation. These motifs would only contain one nitroxide label and could thus contribute to an immobilized *cw*-EPR spectrum, but not to pulsed interaction experiments. As discussed above, the duplex-hairpin equilibrium is challenging to force into complete duplex formation under reasonable conditions. The resulting **RNA^{sc_ext_NO}** distance distribution (Figure 81C, black line) was in excellent accordance to the calculated model system (Figure 81C, shaded in cyan, model shown on the right). The new experimental peak value appeared at 5.49 nm which exactly fits the expected outcome. Successful development of a novel nanometer-range ruler approach for inter-spin measurements in RNA could thus impressively be confirmed. Interestingly, a reduced spin labeling efficiency was observed for **RNA^{sc_ext_NO}** as compared to **RNA^{sc_NO}**, (60 % vs. 99 %, respectively, values were calculated from spin counting to UV absorption ratio). r**TPT3^{NO}** incorporation fidelity might be sensitively depending on the particular sequence employed, however, both values match or even outrange efficiencies of currently available techniques[265,266,275].

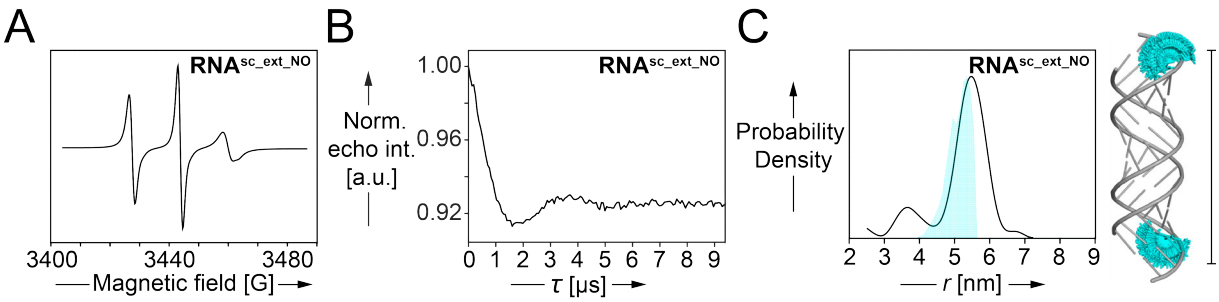


Figure 81. EPR investigations on self-dimerized **RNA^{sc_ext_NO}**; A. *cw*-EPR spectrum; B. Background-corrected and fitted PELDOR time trace (for raw data see section 5.3.18); C. Experimental distance distribution (black line) overlaid with modeled system calculations (cyan shaded, model shown on the right).

Further validation of **RNA^{sc_NO}** and **RNA^{sc_ext_NO}**

To get more information on the duplex RNA sequences, additional experiments should support the EPR-derived findings. Circular dichroism (CD) spectra, UV melting curves, and native PAGE analysis in comparison to canonical sequences originating from commercial solid-phase synthesis should give insight into integrity of double strand formation as well as the impact of r**TPT3^{NO}**:rA mismatches as consequence of the unnatural base modification (Figure 82). The CD spectra exhibit predominant maxima at 260 nm, which is characteristic of antiparallel duplexes.[381] Notably, the 18-mer control sequence **RNA^{sc_n}** consistently exhibited a slightly shifted spectrum under the tested conditions (see 5.3.16), presumably an artifact of extensive hairpin formation.[382]

T_m calculation from the melting curves of the tested r**TPT3^{NO}**-modified and native RNA sequences revealed a lowering of 9.9 °C and 6.2 °C per mismatch introduced by unnatural nucleotide insertion for **RNA^{sc_NO}** and **RNA^{sc_ext_NO}**, respectively (see section 5.3.17). It is conceivable, that the mismatches have a more pronounced impact on the stability of the shorter duplex of **RNA^{sc_NO}**. Decreased melting temperatures of up to 5-10 °C per modification have been reported using spin labels attached to canonical base scaffolds, especially when self-complementary sequences were investigated[127,258,265,383]. The presented novel approach using an UBP for site-directed nitroxide introduction thus profitably adds to the existing methods and is a highly valuable and promising technique.

Additionally, duplex conformations of both **RNA^{sc_NO}** and **RNA^{sc_ext_NO}** were confirmed via native PAGE analysis. Interestingly, the short control sequence **RNA^{sc_n}** showed the monomer or hairpin structure as major component under the applied conditions, supporting the unexpected behavior in CD experiments.

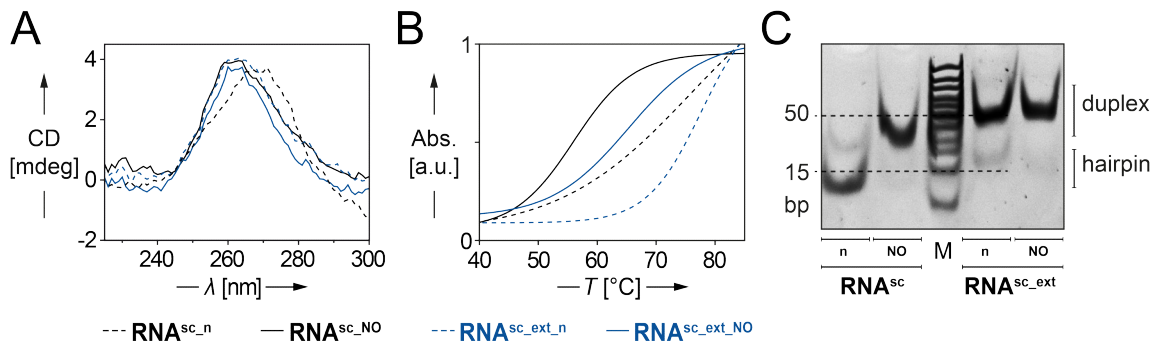


Figure 82. Analysis of self-complementary r**TPT3^{NO}**-modified RNA duplexes **RNA^{sc_NO}** and **RNA^{sc_ext_NO}** in comparison to their unmodified analogs **RNA^{sc_n}** and **RNA^{sc_ext_n}**; A. Averaged CD spectra; B. Fitted UV melting curves; C. 20 % native PAGE (*SYBR Safe* stain).

HPLC-MS analysis of r**TPT3^{NO}**-modified RNAs

Crude (DNase-digested and gel-filtrated) transcripts of the modified self-complementary sequences **RNA^{sc_NO}** and **RNA^{sc_ext_NO}** were further evaluated using HPLC-MS to ensure integrity and proper incorporation of r**TPT3^{NO}** nucleotides (Figure 83). Both constructs could easily be confirmed. In the case of **RNA^{sc_ext_NO}** the truncated transcription product could also be detected; generally higher background signals of this sample do not correlate with oligonucleotide species and are not expected to affect the overall transcription efficiency.

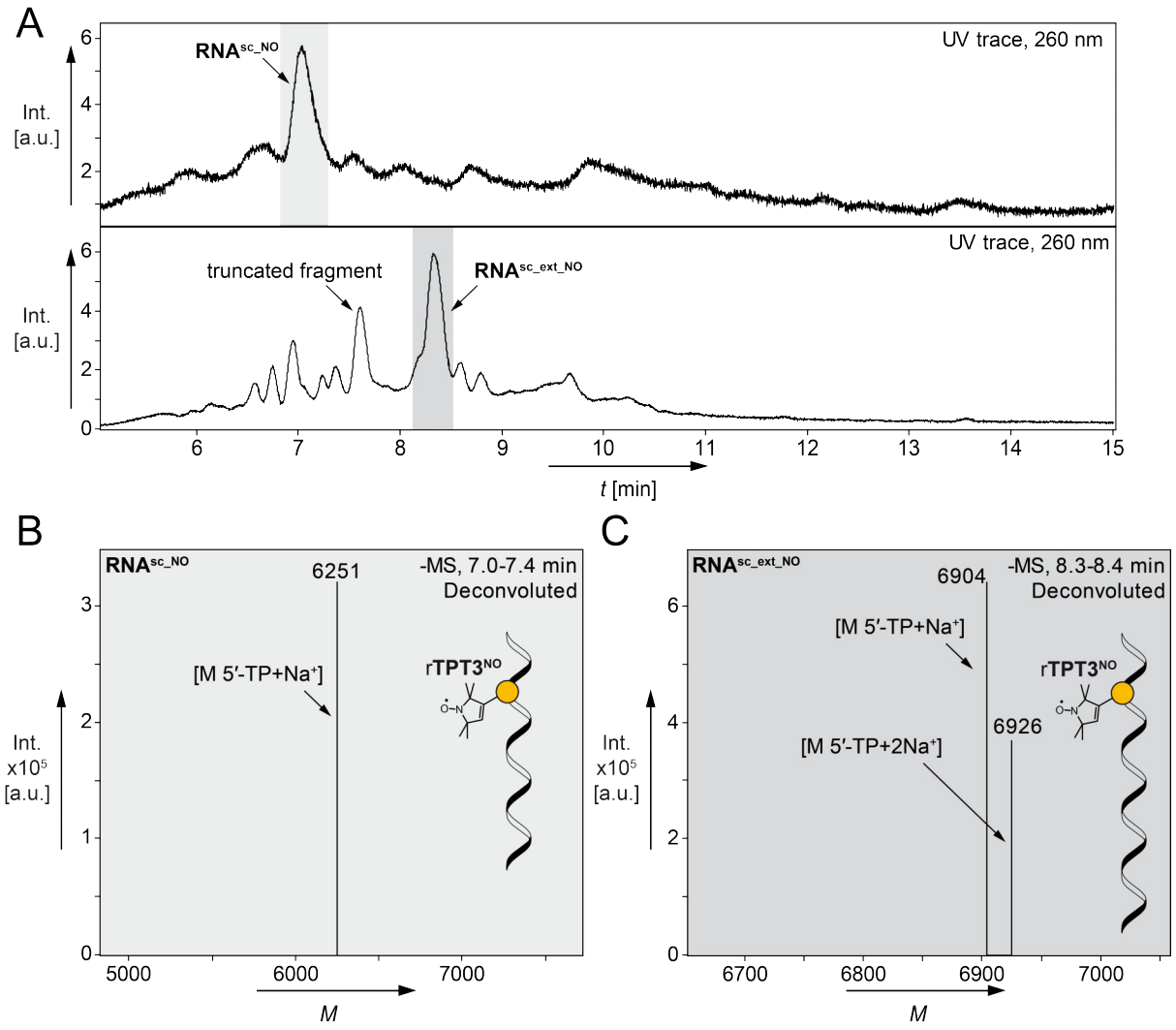


Figure 83. HPLC-MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) of r**TPT3^{NO}**-modified self-complementary transcripts **RNA^{sc_NO}** and **RNA^{sc_ext_NO}**; A. UV traces of crude, DNase-digested and gel-filtrated samples; B. Deconvoluted ESI$^-$ mass spectrum of **RNA^{sc_NO}** ($M_{calcd.}$ = 6252 for M 5′-TP+Na$^+$); C. Deconvoluted ESI$^-$ mass spectrum of **RNA^{sc_ext_NO}** ($M_{calcd.}$ = 6902 for M 5′-TP+Na$^+$, $M_{calcd.}$ = 6924 for M 5′-TP+2Na$^+$).

Incorporation efficiency of r**TPT3^{NO}** TP (**61**) during *in vitro* transcription from short DNA templates

T7 transcriptions from d**NaM**-modified DNA templates in the presence of counterpart r**TPT3**-derived triphosphates result in partial truncation as discussed in section 3.1.2.4. This allows estimation of incorporation efficiency by the ratio of full-length to truncated RNA sequences obtained from band intensities using DPAGE separation. Independent small scale transcriptions were carried out to analyze incorporation behavior of the novel nitroxide-functionalized r**TPT3** nucleotide (Figure 84). Strikingly, transcription from **DNA^{sc_NaM}** with r**TPT3^{NO}** TP (**61**) added to the reaction mix showed ca. 78 % full-length transcript **RNA^{sc_NO}**. Similarly, complete **RNA^{sc_ext_NO}** sequence was yielded with approx. 72 % compared to its truncated fragment. Within the range of accuracy achievable via the presented experimental setup, the incorporation efficiency of both sequences is comparable.

Generally, even higher ratios of full-length transcripts is expected for longer DNA templates since T7 RNA polymerase performs more and more stable transcription when elongating sequences in greater distance to the fragile initiation phase. Therefore, run-off should occur to a lower extent applying this approach for larger construct preparation.

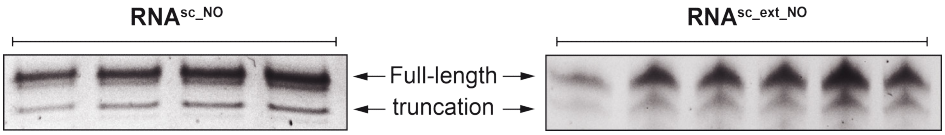


Figure 84. 20 % DPAGE analyses (*SYBR Safe* stain) assessing incorporation efficiency of r**TPT3^{NO}** TP (**61**) into RNA during *in vitro* transcription from **DNA^{sc_NaM}** and **DNA^{sc_ext_NaM}**; full-length transcripts of **RNA^{sc_NO}** and **RNA^{sc_ext_NO}** were obtained with 78 ± 2 % and 72 ± 3 % of total RNA, respectively.

Challenges and limitations of the approach

For short constructs such as **RNA^{sc_NO}** and **RNA^{sc_ext_NO}**, truncated fragment and full-length transcript are resolvable by PAGE and thus separable to facilitate maximum integrity of the sample. This is particularly important when hybridization plays a key role for downstream applications as it is in the case of duplex formation. Presence of sequences other than desired oligonucleotides will hamper proper annealing and eventually give poor results. For longer constructs, however, this might not be feasible. Therefore, caution is advised working with calculated concentrations by UV absorption as the actual amount of desired RNA might be lower. Further uncertainties arising from UV-based concentration determination are the decreased absorption properties of dsRNA compared to ssRNA (75 % based on calculations using the online tool *OligoCalc*[384]) and the unknown extinction coefficient of the r**TPT3^{NO}** nucleobase.

During this study, production of sufficient amounts for PELDOR measurements (≥1 nmol is favorable) was performed several times. Notably, reproduction of the experimental results was occasionally challenging. Electroelution from preparative DPAGE purified samples was suspected to negatively impair the nitroxide spin label. The sterically hindered radical is stable under physiological conditions and in the presence of mild reducing agents such as dithiothreitol (DTT), which is commonly used in transcription reactions. However, increased salt concentration during electroelution (8 M NH_4OAc) and relatively high currents of up to 150 mA might be reasonable for loss of spin activity in some cases. Protocols were changed to a 'crush and soak' elution from DPAGE separated RNAs followed by desalination (see section 5.3.11). Unfortunately, a continuous wave spectrum of the self-complementary **RNA^{sc_ext_NO}** annealed in Mg^{2+}-containing phosphate buffer (see section 5.3.18) indicated very mobile nitroxide labels (Figure 85). Failing removal denaturing urea, which was co-eluted from the DPAGE matrix possibly inhibited proper hybridization of the duplex and PELDOR experiments were omitted due to this moderately promising pre-evaluation.

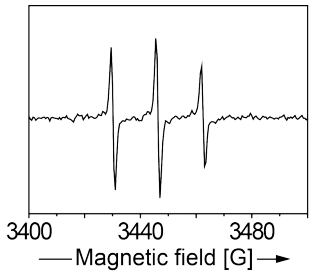


Figure 85. *cw*-EPR spectrum of **RNA^{sc_ext_NO}** after DPAGE separation, desalination and hybridization.

Ongoing studies suggest instability of the r**TPT3**-derived nucleotide due to currently unconfirmed circumstances, which will be discussed below (section 3.2.2.2).

In this work, HPLC purification was determined to be the best purification method. However, *in vitro* transcribed RNAs usually elute in quite broad and partially overlapping peaks. This might require careful and iterative HPLC purifications, which are afflicted with considerable loss of material. Furthermore, separation of truncated from full-length transcripts might be challenging especially for longer sequences. Considering establishment of an optimized HPLC protocol is recommended for prospective experiments involving spin labeled transcripts such as **RNA^{sc_NO}**. High-resolution reversed phase columns, *e.g.* the *XTerra® MS* (2.1×100 mm, 5 µm) by *Waters*, would be a promising system to be tested. For reasons of cost-effectiveness, the ideal conditions could be developed using chemically synthesized RNAs resembling truncated and full-length transcripts.

Notably, some of the recorded spectra exhibited exceptionally steep decay at the lower end of the time trace, indicating rather short inter-spin distances. Unspecific 3′-extension with r**TPT3^{NO}** nucleotides is not expected for such short constructs and no mass differing from

target sequences **RNA$^{sc(_ext)_NO}$** were detected performing HPLC-MS analyses. However, the phenomenon was not precisely excluded. Non-templated 3′-incorporation of an r**TPT3^{NO}** building block would result in a two-base distance to the intended modification site, which would very well account for short distance patterns. *cw*-EPR spectra at cryogenic temperatures could give insights into such issues. Experiments towards this goal are underway.

Employing 2′-OMe-modified nucleotides at the 5′-end of the DNA templates could serve to effectively avoid potential unspecific 3′-elongation of *in vitro* transcripts.[48,368,369] Preliminary attempts gave excellent results (Figure 86) and will lead to successful suppression of undesired effects in future efforts.

In detail, **DNA^{sc_NaM_OMe}** bearing two 2′-OMe-nucleotides at the 5′-end was subjected to T7 *in vitro* transcription in the presence of r**TPT3^{NO}** TP (**61**) (Figure 86A). HPLC-MS analysis of the resulting crude reaction showed two peaks, corresponding to the truncated fragment and full-length **RNA^{sc_NO}** co-eluting with -1 transcript **RNA^{sc_NO}-1**, all of which were also observed by DPAGE (Figure 86B). Identities were further verified by deconvolution of raw ESI$^-$ traces (Figure 86C/D). Indeed, T7 RNA polymerase effectively falls of the template at 2′-OMe-modified nucleotide species, thus non-templated 3′-extension of transcripts is avoided.

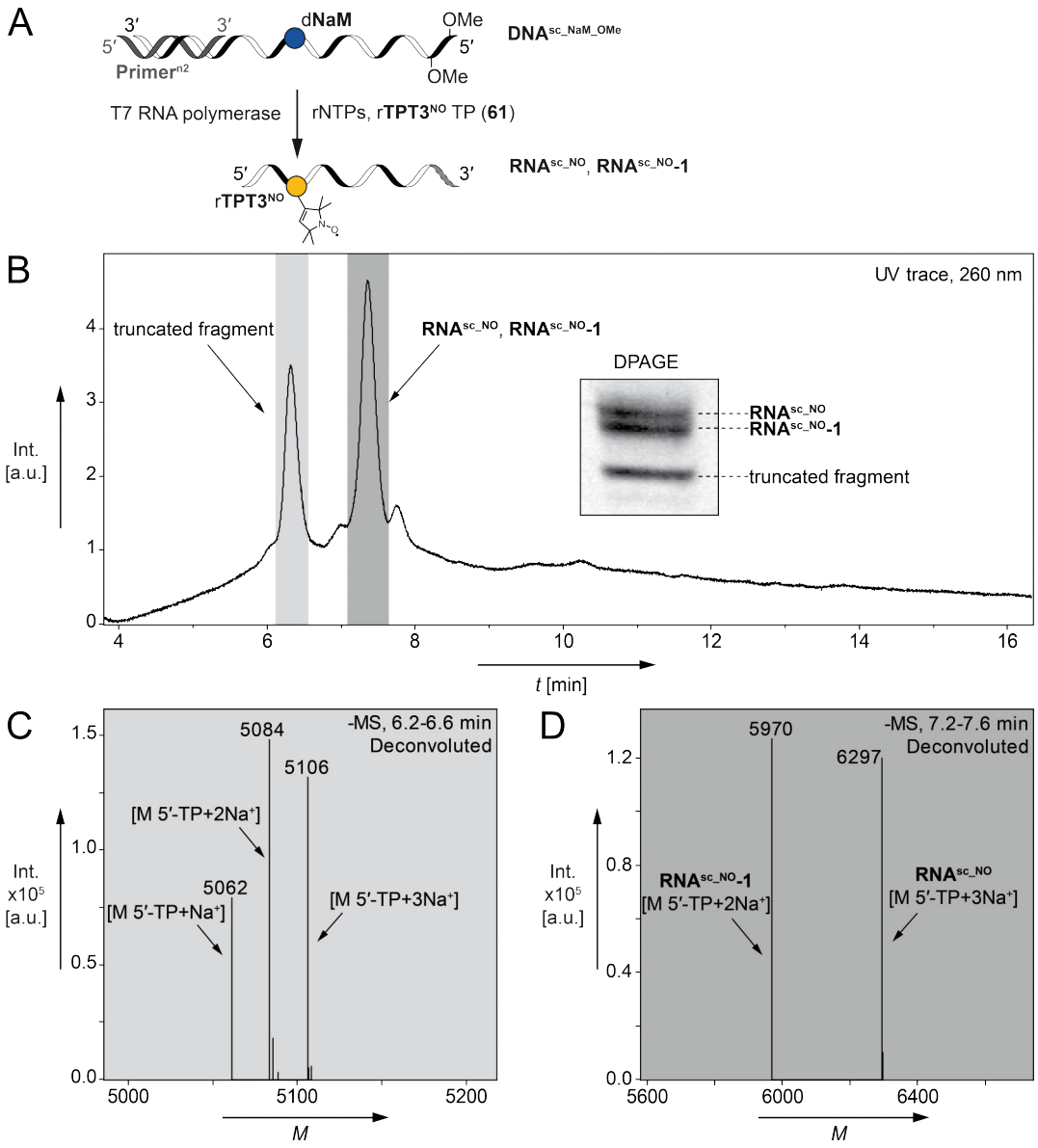


Figure 86. Characterization of a crude DNase-digested and gel-filtrated transcription from **DNA^{sc_NaM_OMe}** in the presence of r**TPT3^{NO}** TP (**61**); A. Schematic representation; B. UV-trace of an HPLC-MS analysis (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) and 20 % DPAGE analysis (insert, *SYBR Safe* stain); C. Deconvoluted ESI$^-$ mass spectrum of the truncated fragment eluting at 6.2-6.6 min ($M_{calcd.}$ = 5062 for [M 5′-TP+Na$^+$], $M_{calcd.}$ = 5084 for [M 5′-TP+2Na$^+$], $M_{calcd.}$ = 5106 for [M 5′-TP+3Na$^+$]); D. Deconvoluted ESI$^-$ mass spectrum of **RNA^{sc_NO}-1** and **RNA^{sc_NO}** eluting at 7.2-7.6 min (**RNA^{sc_NO}-1**: $M_{calcd.}$ = 5969 for [M 5′-TP+2Na$^+$], **RNA^{sc_NO}**: $M_{calcd.}$ = 6297 for [M 5′-TP+3Na$^+$]); experiments performed by J. Mehl.

Remarkably, experiments involving short self-dimerizing RNA sequences generally showed considerable heterogeneity and improvable reproducibility in analytic procedures. This is most likely caused by a sensitive duplex-hairpin equilibrium, which is hardly avoidable.[378,379,381,382,385] Future studies could benefit from complementary sequences that are neither designed nor prone to self-hybridize; therefore, the possibility of alternative structure formation could be avoided, facilitating easy access to consistent data.

In summary, the novel method to site-specifically introduce spin labels into RNA by application of r**TPT3^{NO}** TP (**61**) in template-directed genetic alphabet expansion transcription from d**NaM**-containing DNA presents a significantly advanced procedure. In comparison to *e.g.* the former method involving introduction of r**TPT3^{CP}** nucleotides and subsequent attachment of nitroxides by IEDDA cycloadditions on RNA, this technique is more direct, exhibits less flexibility of spin labels and therefore gives more accurate distance distributions, leading to a valuable nano-scale ruler for investigations of RNA molecules.

Furthermore, the presented technique principally enables preparation of long spin labeled ribooligonucleotides in significant amounts by simple *in vitro* transcription, which would be inaccessible by limits of solid-phase RNA synthesis or ligation approaches. In parallel to the studies on short self-complementary **RNA^{sc_NO}** and **RNA^{sc_ext_NO}** duplexes, further efforts addressed the goal of site-specifically labeling larger sequences with nitroxides, prospectively enabling structural and dynamic insights into long non-coding RNAs. The results will be presented and discussed in section 3.2.2.3.

3.2.2.2 Stability of rTPT3 nucleotides

As mentioned above, consistent reproduction of experimental results using site-specifically r**TPT3^{NO}**-labeled RNAs for PELDOR measurements appeared challenging and several samples prepared using state-of-the-art techniques gave little or no response. During one attempt, an **RNA^{sc_NO}** transcript was purified via a preparative DPAGE, electroelution, and precipitation cascade, followed by HPLC-MS analysis. Unexpected changes in the UV chromatogram were observed (Figure 87A). Comparing the crude transcription reaction (DNase-digested and gel-filtrated) with the DPAGE-purified sample, a minor shift towards lower retention time is visible. Usually, such circumstances should not be concerning since even differences in salt concentrations of the sample sensitively impact apparent retention times. However, the deconvoluted mass spectrum revealed a predominant share with lower mass than the desired **RNA^{sc_NO}** (Figure 87B/C).

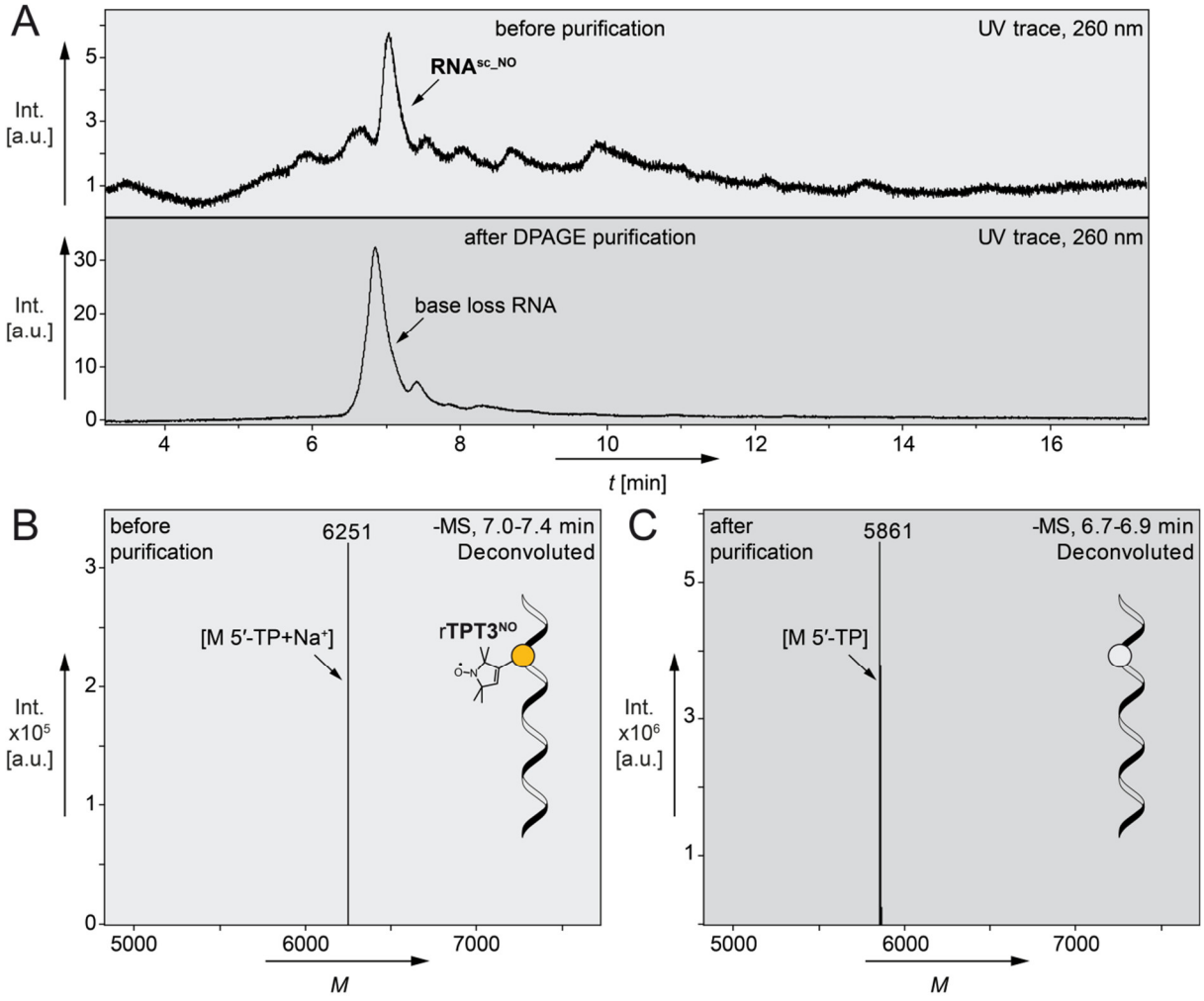


Figure 87. HPLC-MS analysis of self-complementary transcript **RNA^{sc_NO}**; A. UV traces before and after preparative DPAGE purification (see 5.2.4 for details); B. Deconvoluted ESI$^-$ spectrum of intact **RNA^{sc_NO}** ($M_{calcd.}$ = 6252 for [M 5′-TP+Na$^+$]) C. Deconvoluted ESI$^-$ spectrum of base loss **RNA^{sc_NO}** ($M_{calcd.}$ = 5861 for [M 5′-TP-**TPT3^{NO}**]).

Detailed evaluation of this data set led to the suspicion that a loss of the r**TPT3^{NO}** nucleobase analog had occurred. Convincingly, the mass difference that would appear upon substitution of the nucleobase by an amine is ΔM = 369 (Figure 88), which exactly fits the HPLC-MS analysis data (ΔM = 369 in comparison to **RNA^{sc_NO}**). This could explain the difficulties sustainably preparing samples with satisfying modulation depths in connection with nitroxide densities that match the calculated RNA concentration.

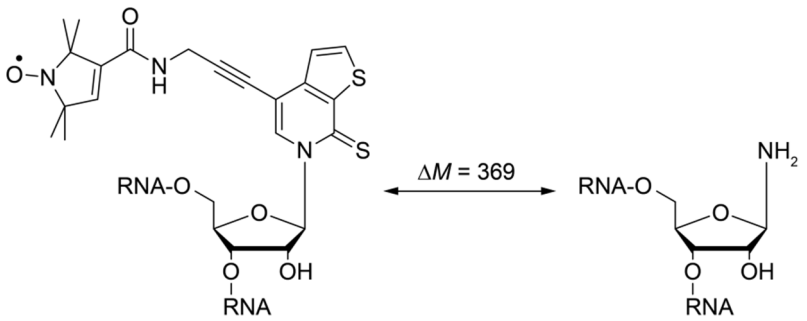


Figure 88. Structural representation of the r**TPT3^{NO}** nucleobase analog loss and mass difference via substitution of the r**TPT3^{NO}** moiety by a free hydroxyl group.

Further supporting this hypothesis, an r**TPT3^{CP}**-modified **RNA^{CP}** transcript was purified via DPAGE, electroelution, and precipitation. HPLC-MS measurement of this sample exhibited a second peak with lower t_R than the RNA of interest (Figure 89A). Similarly, deconvoluted ESI$^-$ data of this peak indeed showed a lower molecular mass (Figure 89B/C). Integration of the UV trace revealed a 35 % share of this species relative to the intact full-length **RNA^{CP}**.

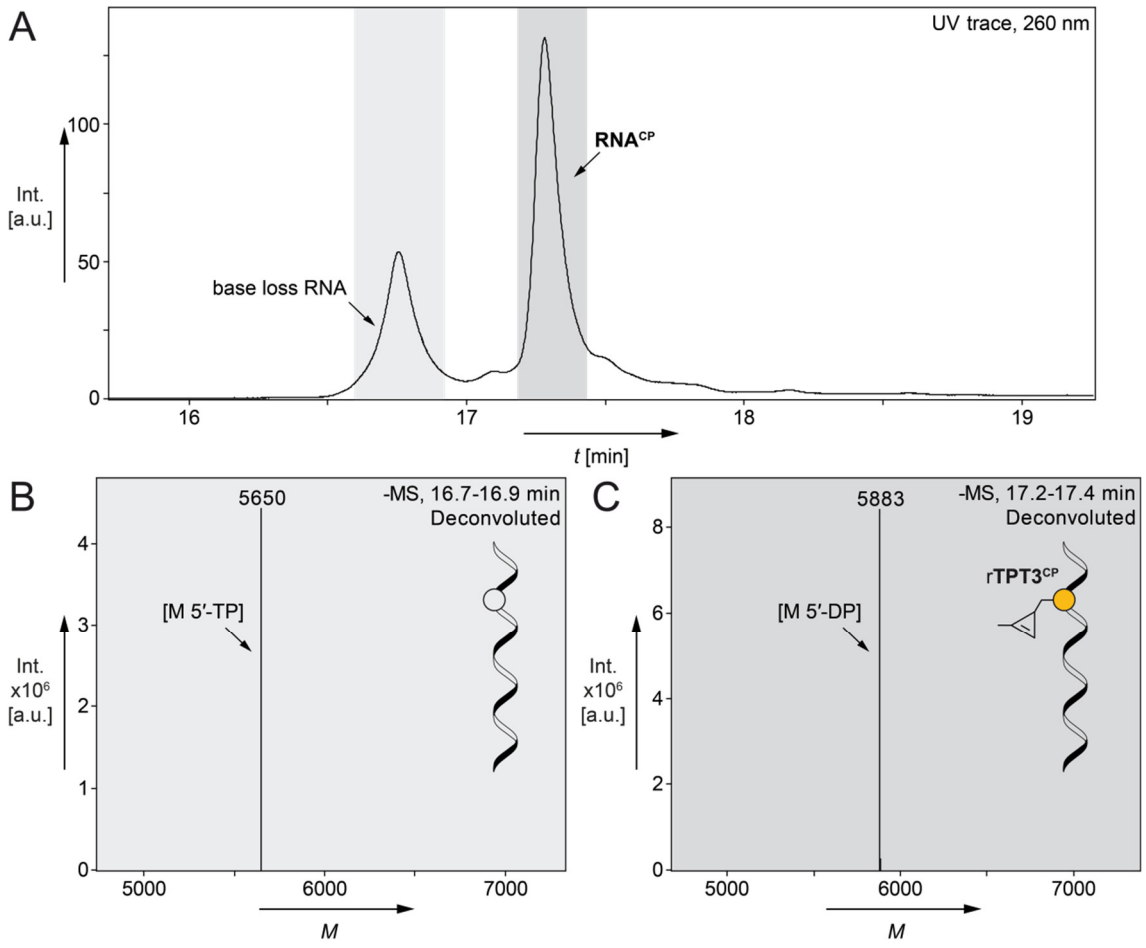


Figure 89. HPLC-MS analysis of preparative DPAGE-purified **RNA^{CP}**; A. UV trace; B. Deconvoluted ESI$^-$ mass spectrum of base loss **RNA^{CP}** eluting at t_R ~16.8 min ($M_{calcd.}$ = 5651 for [M 5′-TP-**TPT3^{CP}**]); C. Deconvoluted ESI$^-$ mass spectrum of intact **RNA^{CP}** eluting at t_R ~17.3 min ($M_{calcd.}$ = 5883 for [M 5′-DP]).

These data support the assumption that base loss events could be of significant abundance. Calculation of the mass difference between an r**TPT3^{CP}**-modified RNA and a 1′-amine-substituted ribose derivative thereof results in Δ*M* = 313, which again is in exact accordance to the peak eluting at ~16.8 min (Δ*M* = 313 in comparison to **RNA^{CP}**).

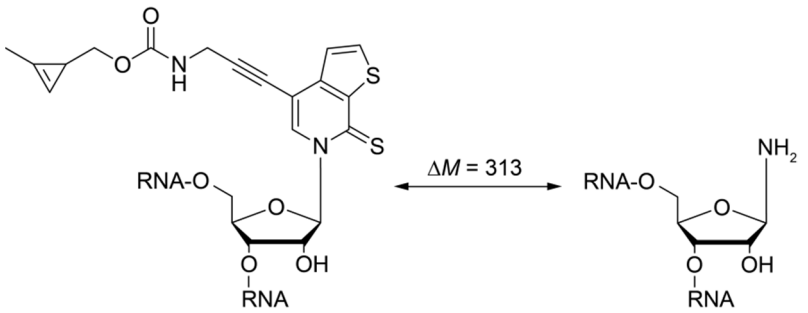


Figure 90. Structural representation r**TPT3^{CP}** nucleobase analog loss and mass difference via substitution of the r**TPT3^{CP}** moiety by a free hydroxyl group.

These independent data clearly suggest that r**TPT3**-derived nucleobase analogs are prone to decomposition using the preparative DPAGE purification approach and several reasons for the base loss to occur can be considered.

Crespo-Hernández as well as Cui and coworkers studied photochemical properties of the **TPT3** and **5SICS** nucleobases[392–395]. These sulfurated heterocycles have been demonstrated (both experimental and computationally) being excitable into long-lasting reactive triplet states by irradiation with near-visible light. Subjecting such a compound to DPAGE, which not only involves elevated voltage and current, but potentially also includes remaining radical species from incomplete polymerization, could have considerably harmful effect on **TPT3**-derived nucleobases.

Furthermore, UV shadowing is applied to visualize RNA before bands of interest are excised for elution, which may further promote photochemical activation.

Additionally, RNA sequences are trapped in highly viscous 8 M ammonium acetate (NH_4OAc) solution during electroelution from PA gels and subsequently precipitated with Ethanol. This process is followed by thorough lyophilization, which might lower the pH and thus favor substitution of the unnatural nucleobase analog by excess ammonia that is formed while evaporating.

Notably, the latter hypothesis of pH-related reactivity would also apply to HPLC purification. Triethylammonium acetate (TEAA) was usually employed as aqueous mobile phase during HPLC. Taking into consideration that NEt_3 has a lower boiling point than acetic acid (~90 °C vs. ~120 °C, respectively), acidification during evaporation of RNA samples, which is commonly applied upon sample preparation, seems reasonable. Implementation of a non-acidifying buffer (*e.g.* triethylammonium bicarbonate) is thus strongly recommended.

Precise and detailed evidence of the particular circumstances under which base loss is supported and to which extent, remains subject of ongoing studies. Due to the free nitroxide radical in the r**TPT3^{NO}** nucleobase, significant reactivity differences in comparison to other **TPT3**-affiliated nucleotides such as r**TPT3^{CP}** can be expected. The presented data suggest a higher risk of base loss for r**TPT3^{NO}** over r**TPT3^{CP}** nucleobase scaffolds.

Involving preparative agarose gel electrophoresis to purify **TPT3^{R}**-containing RNA transcripts is suggested as first course of action. This technique uses lower voltage, shorter time frame, and does not rely on radical polymerization of the matrix. Especially adapting agarose gel electrophoresis to separate small *in vitro* prepared RNAs such as **RNA^{CP}** and **RNA^{sc_NO}** from truncated fragments, careful optimization of experimental protocols is required to enable superior separation properties. Larger transcripts, which do not suffer from extensive run-off transcripts and which could not easily be separated from those, would benefit from subjection to membrane filtration removing free nucleotides and buffer salts. Furthermore, pH-dependent stability measurements of **TPT3** nucleosides and nucleotides employing *e.g.* HPLC-MS could help acquiring insights into the factors that influence base loss.

3.2.2.3 Labeling of the *glmS* ribozyme

As model system testing site-specific spin labeling of longer ribonucleic acids by *in vitro* transcription the *B. subtilis glmS* ribozyme was chosen. Easy accessibility of the construct was proposed based on former studies within this work generating r**TPT3^{CP}**-modified RNA **glmS_RNA^{CP}** (see section 3.1.2.2). Several attempts were made to obtain reliable and reproducible data from nitroxide-functionalized *glmS* sequences. Notably, satisfying performance of respective experiments appeared to be challenging. Many obstacles have been faced, addressed, and overcome to enable straight-forward success in prospective approaches. The following section will chronologically present and discuss the significant stages of research on the project.

<u>First attempts introducing r**TPT3^{NO}** modifications into loop regions</u>

The in-hand dsDNA template prepared by two-step assembly PCR (section 3.1.2.2) was submitted to a T7 *in vitro* transcription in the presence of r**TPT3^{NO}** TP (**61**) yielding doubly nitroxide-labeled **glmS_RNA^{NO}** (Figure 91A) in an analogous manner to **glmS_RNA^{CP}**. After DNase digestion and gel purification, presence and cleavage activity of the construct was confirmed by DPAGE analysis as qualitative measure of correct folding (Figure 91B).

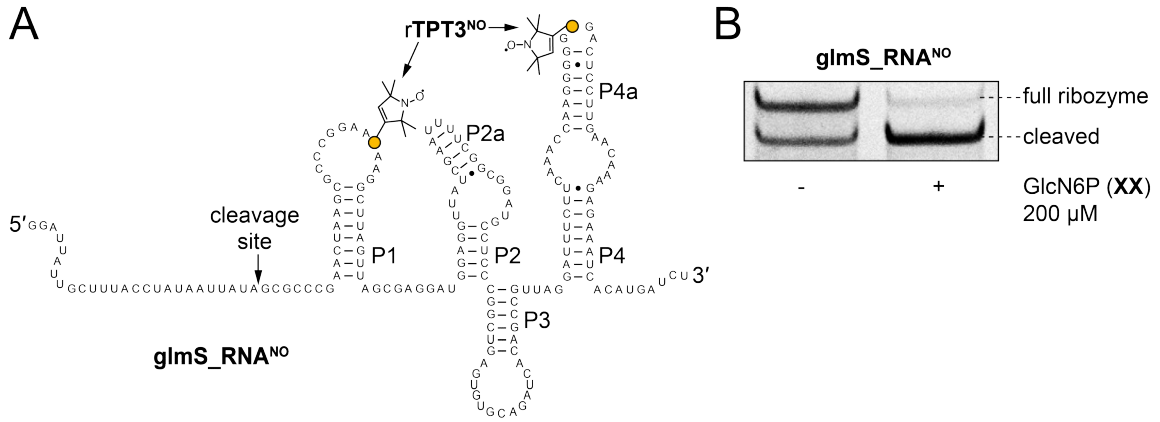


Figure 91. r**TPT3^{NO}**-modified *glmS* construct **glmS_RNA^{NO}** with two nitroxide labels positioned in loop regions of helices P1 and P4a, respectively; A. Schematic ribozyme secondary structure with modification sites; B. 12 % DPAGE analysis (*SYBR Safe* stain) of gel-filtrated **glmS_RNA^{NO}** transcript before and after incubation with cofactor GlcN6P (**1**).

A 2.0 nmol PELDOR sample was prepared according to established protocols (section 5.3.18), which resulted in rather unsatisfactory data such as a mobile nitroxide signal, which have thus been excluded from further investigations.

Unnatural base sites were positioned in loop regions of the **glmS_RNA^{NO}** RNA construct to avoid mismatches and associated destabilization of stem loops. The labeling loci of the *glmS* sequence were expected to be closely approximated to each other and stably embedded into the overall structure in accordance to available *glmS* crystal structures.[70,71,73] However, loop regions bear a remaining risk of nitroxide labels being too flexible for accurate determination of inter-spin distances.

Gel filtration via spin columns was utilized for purification of the EPR sample, which might not quantitatively hold back the large excess of r**TPT3^{NO}** TP (**61**) supplied for *in vitro* transcritpion. Presence of **61** in a PELDOR sample would also result in the increased appearance of a mobile continuous wave signal. Thus, a **glmS_RNA^{NO}** sample was purified by reverse phase HPLC (see 5.3.11 for details) to truthfully remove any r**TPT3^{NO}** TP (**61**) from the transcript. However, *cw*-EPR spectroscopic analysis revealed considerably mobile spin label species.

These pioneering experiments led to the conclusion that spin labeling of RNA in loop regions of this particular construct is unsuitable for PELDOR experiments aiming to determine intramolecular distances in the highly folded *glmS* structure.

glmS ribozyme constructs from fully solid phase-synthesized templates

Fast access to site-specifically spin labeled *glmS* ribozymes via genetic alphabet expansion transcription was aspired using 3′- and 5′-shortened sequences of **glmS_RNA^{NO}**. Template DNAs could be entirely prepared by custom DNA solid-phase synthesis (179 nt, *Ella Biotech*, Martinsried, Germany), thus 5-letter PCR assembly was not needed in this case. Consequently, straight-forward transcription reactions with r**TPT3^{NO}** TP (**61**) and exclusion of instable d**TPT3**:d**NaM** UBP propagation during fusion PCR were simultaneously ensured. d**NaM**-containing antisense/reverse DNA templates **glmS_DNA^{NaM_P1P4a_short_RV}** or **glmS_DNA^{NaM_P1P4L_short_RV}**, each hybridized to an unmodified T7 promoter sequence (**glmS_Primer_FW^{short}**), enabled *in vitro* expanded genetic alphabet transcription employing r**TPT3^{NO}** TP (**61**) to obtain 162 nt transcripts **glmS_RNA^{NO_P1P4a_short}** and **glmS_RNA^{NO_P1P4L_short}**, respectively (Figure 92A).

glmS_RNA^{NO_P1P4a_short} contained an r**TPT3^{NO}** nucleotide in the stem regions of helix P1 and P4a. **glmS_RNA^{NO_P1P4L_short}** is modified at the same position in P1, however, the second nitroxide label was introduced into stem region P4. Furthermore, **glmS_DNA^{n_short_RV}** consisting of canonical bases served to prepare the native RNA sequence in a comparative manner. Introduction of r**TPT3^{NO}** entities into stems of the proposed *glmS* constructs should enable certain fixation of the nitroxide moieties and thus increase the option to obtain distinct inter-spin distances. Introduction of the cognate r**NaM** unnatural base into RNA by *in vitro* transcription in a quantitative manner has not yet been fully established and verified, thus mismatches in the modified helices had to be accepted.

Preparative DPAGE was employed purifying the transcribed RNA to overcome incomplete separation from free triphosphates. Since electroelution was suspected causing occasional inactivation of the spin label (see section 3.2.2.2), a ‘crush and soak’ protocol was established to elute nucleic acids from the polyacrylamide gel matrix followed by desalination of the samples (see section 5.3.11).

Analysis of the purified RNAs via DPAGE clearly shows successful and efficient preparation of the unmodified construct **glmS_RNA^{n_short}** (Figure 92B). In contrast, **glmS_RNA^{NO_P1P4a_short}** and **glmS_RNA^{NO_P1P4L_short}** were yielded in decreased amounts and rather diffuse bands allowed for interpretation of considerable transcript heterogeneity. Detection of cleaved fragments for the modified RNAs could not unambiguously be verified, possibly suggesting folding difficulties.

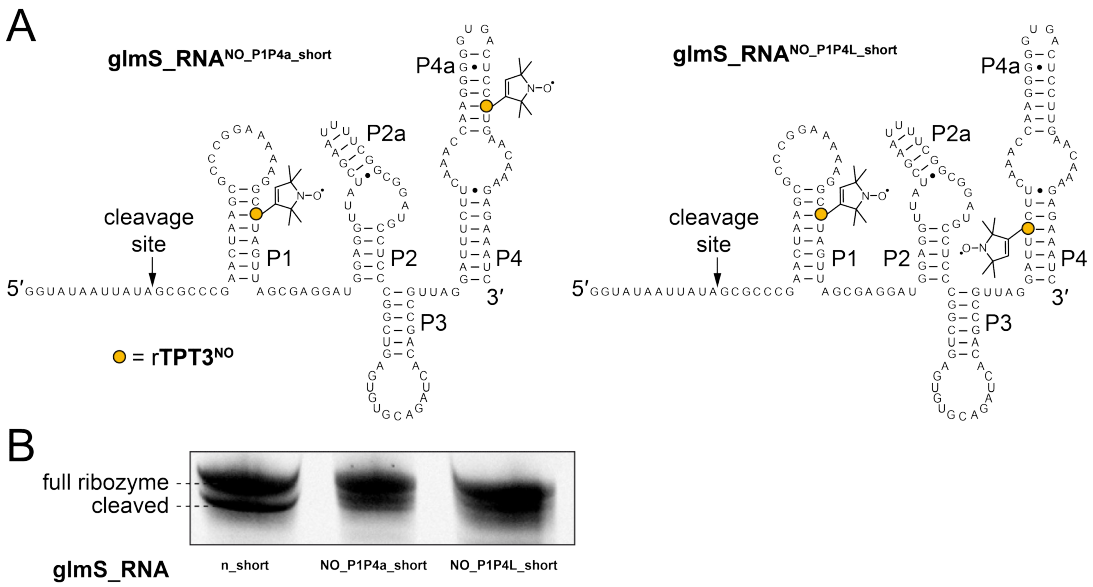


Figure 92. r**TPT3^{NO}**-containing RNA constructs **glmS_RNA^{NO_P1P4a_short}** and **glmS_RNA^{NO_P1P4L_short}** with labels positioned in stem regions P1/P4a and P1/P4, respectively; A. Ribozyme sequences and secondary structures with indicated labeling positions; B. 12 % DPAGE analysis (*SYBR Safe* stain) of DPAGE/crush and soak-purified transcripts in comparison with the unmodified RNA **glmS_RNA^{n_short}**.

Sufficient amounts of **glmS_RNA^{NO_P1P4a_short}** (490 pmol) and **glmS_RNA^{NO_P1P4L_short}** (300 pmol) were prepared via *in vitro* transcription, purified, and subjected to EPR spectroscopic analysis.

Both constructs exhibited comparable results (Figure 93). *cw*-EPR analyses suggested considerable mobility of the spin labels (Figure 93A/D). In combination with rather low modulation (Figure 93B/E), PELDOR experiments (Figure 93C/F) resulted in broad distance distributions that did not allow for gathering precise information.

Several factors could be hampering this approach. Since the particular *B. subtilis glmS* sequence has not been crystallized yet, it remains unknown whether helix P1 and P4/P4a are located in the same reasonable approximation as it is the case for the *B. anthracis* species.[74] Decomposition of r**TPT3^{NO}** nucleotides during preparative PAGE might have occurred (see section 3.2.2.2). Additionally, correct folding of **glmS_RNA^{NO_P1P4a_short}** and **glmS_RNA^{NO_P1P4L_short}** constructs could be hindered by excess urea originating from the gel matrix as discussed for short self-complementary duplexes modified with spin labels (section 3.2.2.1). A broad distribution of misfolded ribozyme states would likely result in such a diffuse distance pattern and act as limiting factor rigidly installing spin labels in these molecules.

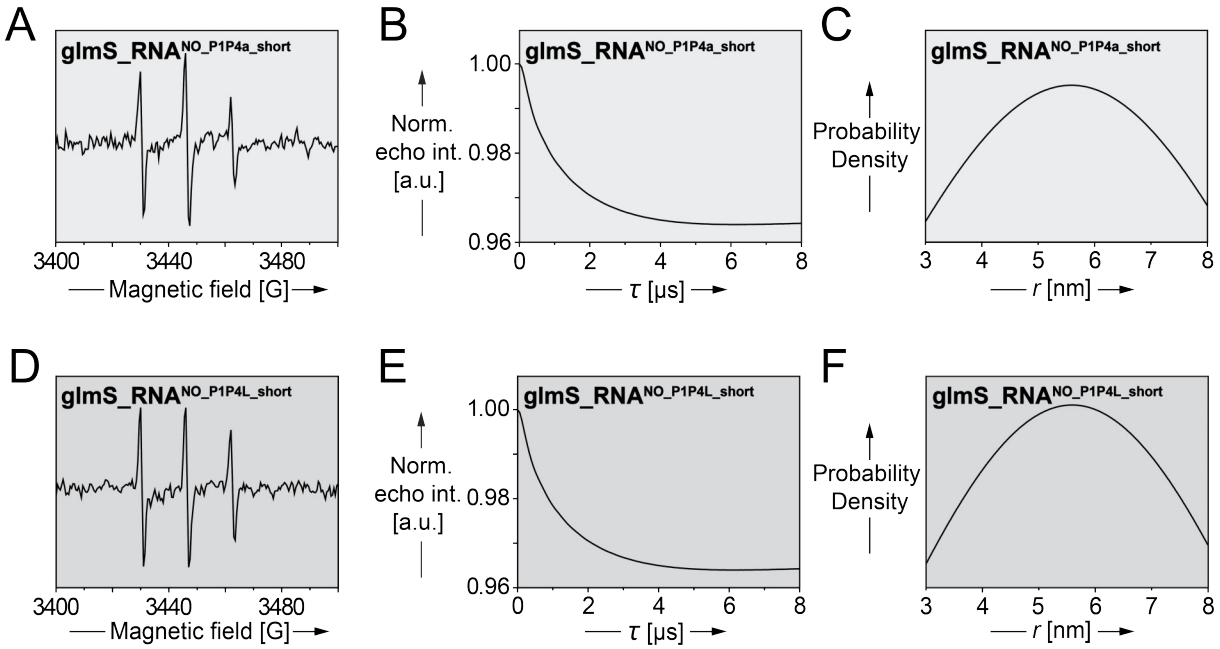


Figure 93. EPR spectroscopic analysis of **glmS_RNA^{NO_P4P4a_short}** (A-C) and **glmS_RNA^{NO_P4P4L_short}** (D-F); A./D. *cw*-EPR spectra; B./E. Fitted background-corrected PELDOR echo time traces (for raw data see section 5.3.18); C./F. PELDOR-derived distance distributions, fitted with Gaussian model.

HPLC purification of *in vitro* transcribed samples yielding **glmS_RNA^{NO_P1P4a_short}** and **glmS_RNA^{NO_P1P4L_short}** was performed to avoid denaturing conditions. Analysis by *cw*-EPR again gave sharp peaks, suggesting very mobile nitroxide species (Figure 94).

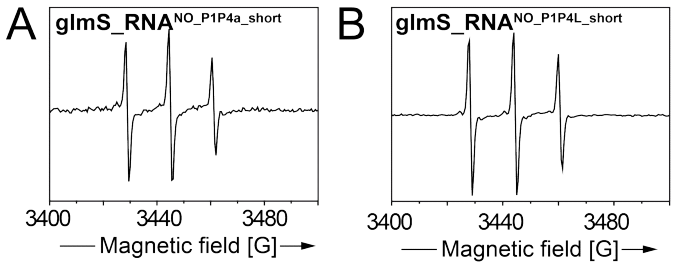


Figure 94. *cw*-EPR spectroscopic analysis of HPLC-purified transcripts; A. **glmS_RNA^{NO_P1P4a_short}**; B. **glmS_RNA^{NO_P1P4L_short}**.

In order to support proper folding of the ribozymes, conditions were subsequently adjusted to higher salt content (final concentrations of 470 mM NaCl and 18 mM $MgCl_2$) followed by re-annealing, which did not lead to increased immobilization of the spin labels.

Preparation of highly structured RNA molecules by *in vitro* transcription from larger DNA templates applying site-specific r**TPT3^{NO}** modifications remained challenging. Full-length synthetic DNA of such length (179 nt) approach the limits of chemical accessibility (*biomers.net, sigmaaldrich.com,* accessed 2019/07). Product sequences are expensive, low yielding, and heterogeneity is strongly increased since truncated fragments that inevitably arise during the synthesis cannot be removed during purification of target nucleic acids.

Therefore, the pronounced goal to employ inter-spin measurements by means of EPR spectroscopy in highly structured RNAs was continued preparing the template DNAs by three-piece 6-letter fusion PCR as described in section 3.1.2.2.

Third-generation templates for spin labeled ribozymes by fusion PCR

Relying on the well-tested and functioning assembly PCR approach, the proposed constructs were changed with respect to the former r**TPT3^{CP}**-modified *glmS* constructs. In accordance to these, two new target sequences were envisaged, which comprised 185 nt and differed in one of two labeling positions.

GlmS_RNA^{NO_P1P4a} (Figure 95A) represents an extended analog to **glmS_RNA^{NO_P1P4a_short}** shown above. r**TPT3^{NO}** nucleotides were introduced in stem regions of P1 and P4a. A crude, gel-filtrated **glmS_RNA^{NO_P1P4a}** transcript analyzed on DPAGE confirmed purity and integrity of the ribozyme.

Additionally, a second construct was prepared. **glmS_RNA^{NO_P4RP4a}** was designed bearing nitroxide labels in helices P4 and P4a, respectively (Figure 95B). This assembly was chosen to ensure close and defined proximity of TEMPYO moieties, facilitating accurate measurement of inter-spin distances. Notably, relative to the shorter variant **glmS_RNA^{NO_P1P4L_short}** used earlier, the P4 labeling position is located opposite in the helix, close to the 3′-end of the ribozyme. DPAGE analysis of a crude **glmS_RNA^{NO_P4RP4a}** transcription reaction, desalted via gel filtration, verified presence and identity of the desired RNA sequence.

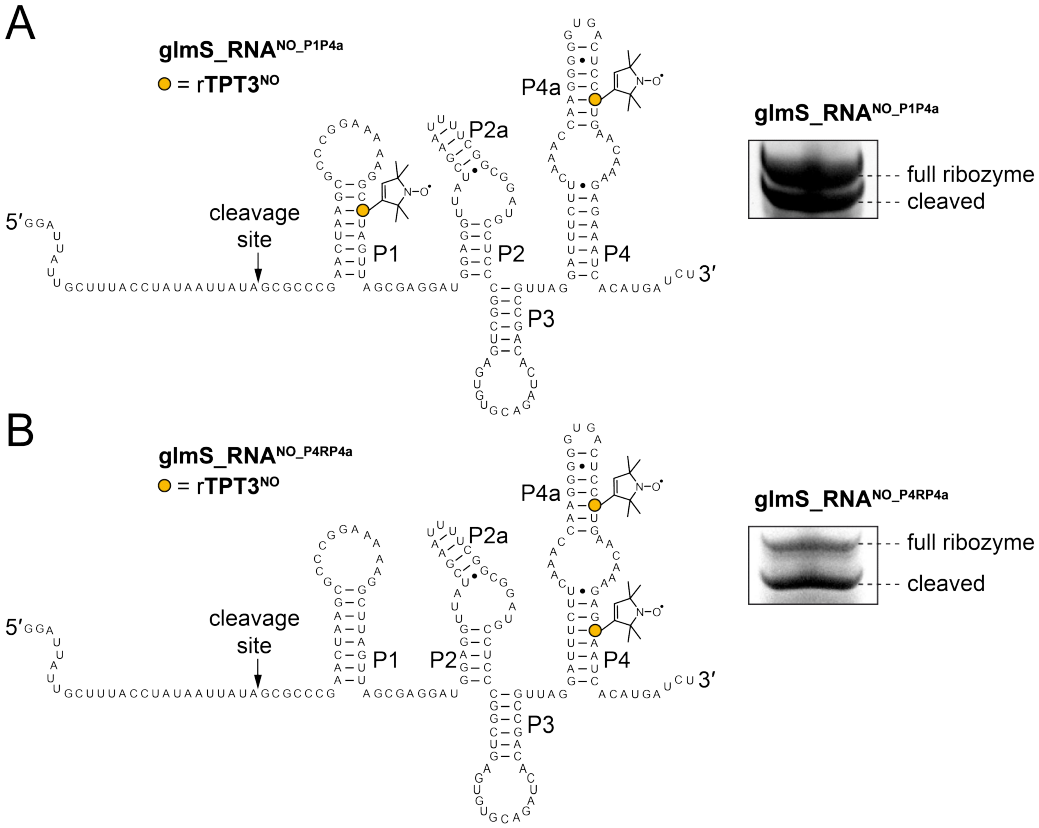


Figure 95. Spin labeled *glmS* constructs **glms_RNA^{NO_P1P4a}** (A) and **glms_RNA^{NO_P4RP4a}** (B) prepared by transcription from fusion PCR-derived DNA templates; Sequences and secondary structures are shown left, 12 % DPAGE analyses of crude, DNase-digested and gel-filtrated RNA samples on the right-hand side confirm successful generation of cleavable ribozymes.

Both ribozyme variants were prepared for PELDOR experiments (**glmS_RNA^{NO_P1P4a}**: 830 pmol, **glmS_RNA^{NO_P4RP4a}**: 1.1 nmol). To accelerate the overall process and to avoid loss of precious material, simple gel filtration was applied while charging decreased volume per spin column to ensure complete absorption of excess r**TPT3^{NO}** TP (**61**).

The apparent spin labeling efficiency of **glmS_RNA^{NO_P1P4a}** was determined to be 115 % by comparison of UV absorption and spin counting (c_{A260} = 32.8 µM theoretical radical species versus c_{EPR} = 37.8 µM nitroxide concentration). Despite this fact, EPR-spectroscopic analysis resulted in low modulation, which did not correlate to an evaluable distance distribution. In combination with earlier efforts to investigate site-specifically nitroxide labeled *glmS* ribozymes including an r**TPT3^{NO}**-modification in stem P1, this particular construct was refrained from further EPR-spectroscopic investigations.

From PELDOR measurements of **glmS_RNA^{NO_P4RP4a}** a full set of interpretable data was obtained. The continuous wave EPR spectrum of the sample (Figure 96A) revealed immobilized nitroxide species. Sufficient modulation was observed in the corresponding echo time trace (Figure 96B) and the PELDOR distance distribution (Figure 96C) exhibited a distinct maximum signal at 4.5 nm as well as a minor peak at ~7.2 nm.

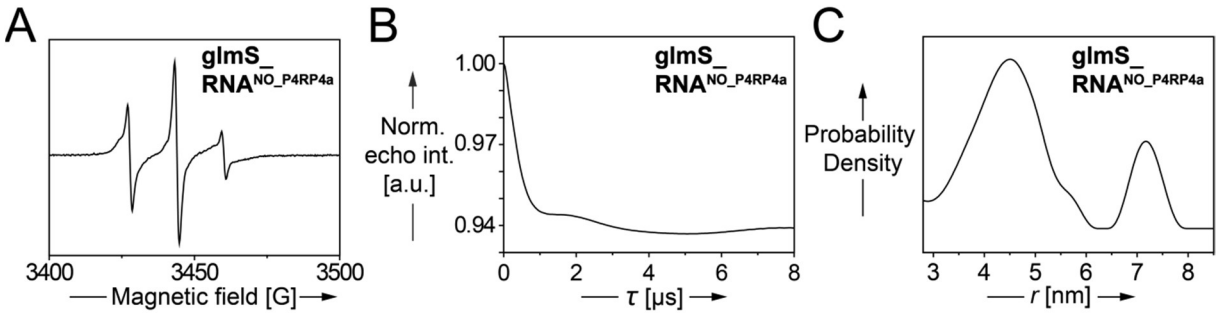


Figure 96. EPR measurements of *in vitro* transcribed, gel-filtrated **glmS_RNA^{NO_P4RP4a}** ribozyme; A. *cw*-EPR spectrum; B. Background-corrected PELDOR echo time trace (for raw data see section 5.3.18); C. PELDOR-derived distance distribution.

The shape of the third peak in the *cw*-EPR spectrum at approx. 3460 G indicates presence of two or more spin label species with different immobilization. Thus, the observed spectrum might depict a superposition of several signals originating from differently folded ribozyme states.

Regarding the respective spin labeling efficiencies of this *glmS* variant, significant inconsistencies were observed. RNA concentration determined by optical density was c_{A260} = 35.2 µM (corresponding to 70.4 µM spins), whereas spin counting yielded c_{EPR} = 112.0 µM nitroxides, which results a nominal 159 % labeling efficiency.

Notably, RNA concentrations were determined using the online tool *OligoCalc*[384] in the ssRNA mode from the apparent A_{260} values employing canonical sequences as input. As discussed for self-complementary spin labeled duplexes (section 3.2.2.1), considerable differences arise whether ssRNA or dsRNA is used as calculation basis. The investigated *glmS* ribozymes naturally possess both single and double stranded regions; therefore, UV absorption values might be associated with significant deviation. On the other hand, an uncertainty of ca. 10 % is considered inherent to spin counting.

Highly efficient labeling would suggest increased modulation, which does not correspond to the apparent PELDOR echo time trace.

Experiments employing short self-complementary RNA duplexes (section 3.2.2.1) revealed at least 70 % incorporation efficiency of single r**TPT3^{NO}** nucleotides during *in vitro* transcription from d**NaM**-containing DNA templates. Transferring these findings to the presented approach introducing two labels into *glmS* ribozymes, a theoretical maximum of 49 % doubly modified can be considered. A certain, yet unaddressed share of UBP loss during 6-letter PCR assembly and amplification of the corresponding dsDNA template might also account for a decreased amount of doubly spin labeled **glmS_RNA^{NO_P4RP4a}**. Furthermore, misfolding of the generated transcript could be a significant factor explaining the experimental outcome.

Taking possible sample impurities with *e.g.* inseparable truncated fragments into consideration, quantitative labeling is not assumed for **glmS_RNA^{NO_P1P4a}** and

glmS_RNA^{NO_P4RP4a}. An unspecific 3′-extension event involving r**TPT3^{NO}** TP (**61**) was suspected to be the most plausible origin for over-quantitative labeling efficiency. Moreover, this could also explain the second peak observed in the distance distribution of **glmS_RNA^{NO_P4RP4a}** at ~7 nm and was thus subject of investigations that will be discussed below.

As mentioned above, the particular *B. subtilis* RNA sequence employed in this study has not been analyzed by crystallographic methods yet. Due to extensive conservation of the chosen P4/P4a stems in comparison to the *B. anthracis glmS* construct, great similarity can be expected (Figure 97A).[61,73] Arguably, the loop region between stem regions P4 and P4a within the *B. subtilis* sequence suggests a slightly longer distance between the respective spin label position sites. r**TPT3^{NO}** nucleotides were placed at the sites of interest into a *B. anthracis glmS* ribozyme crystal structure[74] by *MtsslWizard*[380] simulation (Figure 97B). An overlay of the model distance region with PELDOR probability density data showed significant overlap (Figure 97C), which confirms the expectations of a most probable distance at 4.5 nm.

With this, a successful first approach to measure inter-spin distances from site-specifically introduced nitroxide labels via *in vitro* transcription could be demonstrated.

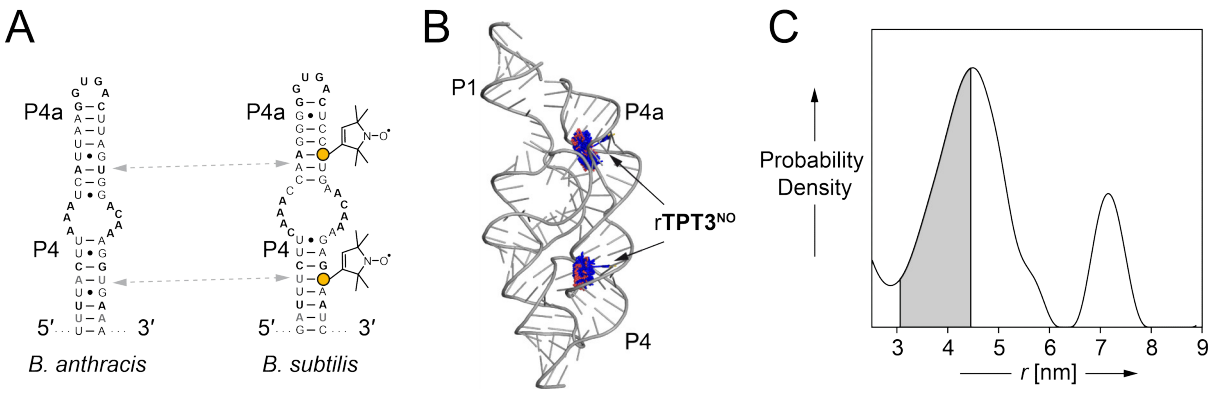


Figure 97. glmS_RNA^{NO_P4RP4a} structural analysis; A. Comparison of P4/P4a helices from *B. anthracis* and *B. subtilis glmS* ribozymes[61,73], conserved nucleotides are indicated in bold black, switched base pairing is shown by bold gray letters; B. *MtsslWizard*[380] simulation of r**TPT3^{NO}** modifications in helix P4 and P4a modeled in the crystal structure of *B. anthracis glmS* ribozyme (PDB file: 3l3c[74]); C. PELDOR-derived distance distribution (black line) in overlay with the area obtained from the model structure (shaded in gray).

It is widely accepted, that *glmS* ribozymes fold into their proper three-dimensional structure independently of GlcN6P (**1**) presence.[54,73,74] Providing further support for this issue was intended using the presented method. The **glmS_RNA^{NO_P4RP4a}** PELDOR sample was supplemented with cofactor **1** to a final concentration of 200 µM, heated to 95 °C for 2 min, and incubated at room temperature overnight to assure completion of the cleavage reactions. The echo time trace was not affected by this approach, confirming the general observation of

a preformed GlcN6P (**1**) binding site. Upon recollection of the EPR RNA sample and subsequent desalination, **glmS_RNA^{NO_P4RP4a}** was analyzed via HPLC-MS and DPAGE. The 5′-cleaved fragment eluting at ~8 min (Figure 98A) was identified via HPLC-MS bearing a 2′,3′-cyclic phosphate (Figure 98B) in accordance to the proposed cleavage mechanism[61] (see section 1.2.1). The spin labeled ribozyme construct showed predominant presence of the cleaved species upon DPAGE separation (insert in Figure 98A). This positive cleavage test confirms correct folding and associated catalytic activity of **glmS_RNA^{NO_P4RP4a}**.

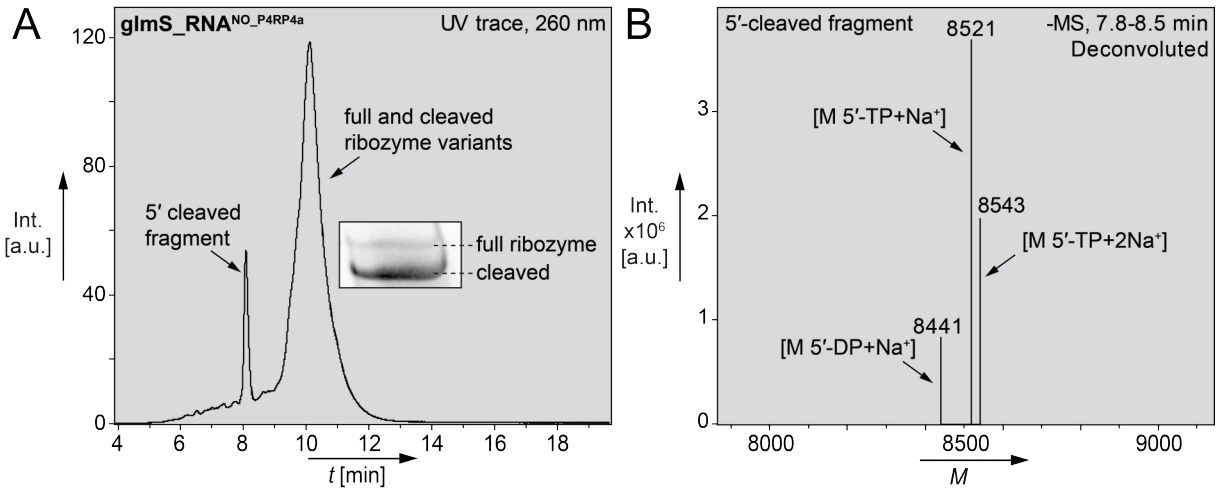


Figure 98. HPLC-MS (3→20 % MeCN/buffer in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.3.13) and 12 % DPAGE analyses of r**TPT3^{NO}**-modified *glmS* construct **glmS_RNA^{NO_P4RP4a}**; A. UV trace; B. Deconvoluted ESI$^-$ mass spectrum of the 5′-cleaved fragment eluting at ~8 min ($M_{calcd.}$ = 8441 for [M 5′-DP-3′-cP +Na$^+$], $M_{calcd.}$ = 8521 for [M 5′-TP-3′-cP+Na$^+$], $M_{calcd.}$ = 8543 for [M 5′-TP-3′-cP+2Na$^+$].

Simulation of the nitroxide-modified ribozyme **glmS_RNA^{NO_P4RP4a}**

Structural investigations on biomolecules can be supported and predicted using molecular dynamics (MD) calculations.[48,97,98,275] S. Kath-Schorr simulated the dynamics of **glmS_RNA^{NO_P4RP4a}** containing two r**TPT3^{NO}** nucleotides at the proposed positions in stems P4 and P4a as well as an additional r**TPT3^{NO}** entry at the 3′-end of the sequence (Figure 99A) representing an unspecific incorporation event. The starting geometry input of the arising 186 nt construct was derived from an *B. anthracis* ribozyme X-ray crystal structure.[74] All necessary nucleotide mutations and extensions to resemble identical sequence context of the *B. subtilis* **glmS_RNA^{NO_P4RP4a}** ribozyme[61] functionalized with r**TPT3^{NO}** nucleotides in helices P4/P4a and at the 3′-terminus were included. Calculations (300 K) were performed with a 72 ns equilibration period followed by 60 ns production time using the *GROMACS*[99,100] package equipped with the *CHARMM36*[101,386] (2018/01) all-atom force field set.

Strikingly, the main distance at 4.5 nm found by PELDOR (Figure 99B) overlaps well with the MD simulation for the P4-P4a distance (d(P4-P4a), light blue, ca. 4–6 nm, peak at ~5.0 nm)

verifying the measurement data in a comforting manner. The approximation of the P4 r**TPT3^{NO}** position to the 3′-end, however, is also in excellent agreement to this range (d(3′-P4), dark blue, ca. 3.5–5.0 nm, peak at ~4.5 nm). Therefore, superposition of both distances resulting in broadening of the peak could be considered. Furthermore, d(3′-P4a) (red, ca. 7.0–9.5 nm, peak at ~8.5 nm) representing the distance of the P4a r**TPT3^{NO}** label to the 3′-terminus of the ribozyme is in quite good accordance to the PELDOR distance of ~7.0–8.0 nm. Parameters in these PELDOR measurements were not optimized to higher distance values as they have not been expected.

In summary, MD data strongly support the experimental outcome of the distance distribution obtained by PELDOR and prove general applicability of site-specific r**TPT3^{NO}** introduction into larger and highly structured RNA sequences to serve as a nanometer-range ruler. Moreover, computational calculations gave rise for investigations regarding the longer experimental distance by means of non-templated 3′-extension during *in vitro* transcription. Attempts to gain insights on this issue are discussed in the following.

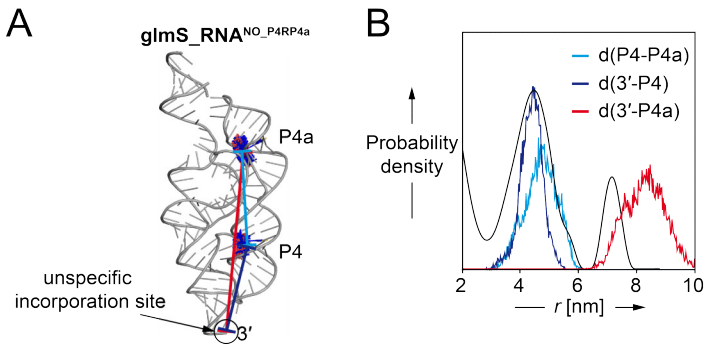


Figure 99. Intramolecular distances in the r**TPT3^{NO}**-labeled ribozyme **glmS_RNA^{NO_P4RP4a}**; A. Distances of spin labels schematically modeled into the crystal structure of *B. anthracis glmS* ribozyme (PDB file: 3l3c[74]), shown are intended nitroxide labels in helix P4 and P4a as well as the 3′-end representing unspecific incorporation of r**TPT3^{NO}** TP (**61**); B. Overlay of PELDOR distance distribution and MD calculations.

A *glmS*-Hammerhead (HH) ribozyme construct

Since the peak appearing at larger values of the **glmS_RNA^{NO_P4RP4a}** PELDOR spectrum could be related to unspecific 3′-labeling, the topic should be addressed by generating a ribozyme variant with a defined sequence end. Therefore, a *glmS* construct flanked by the hammerhead (HH) self-cleaving RNA, **glmS_RNA^{NO_P4RP4a_HH}**, was aspired (Figure 100). The HH sequence is known to efficiently perform Mg^{2+}-dependent cleavage directly upstream of its 5′-end and should proceed smoothly in a co-transcriptional manner. This design would release the *glmS* construct with a homogenous 2′,3′-cyclic phosphate end.[48,89,387–389] Unspecifically appended r**TPT3^{NO}** nucleotides could then be separated from the desired RNA along with the cleaved HH fragment. If the hypothesis is valid, this should ultimately result in a PELDOR spectrum deficient of the high distance peak.

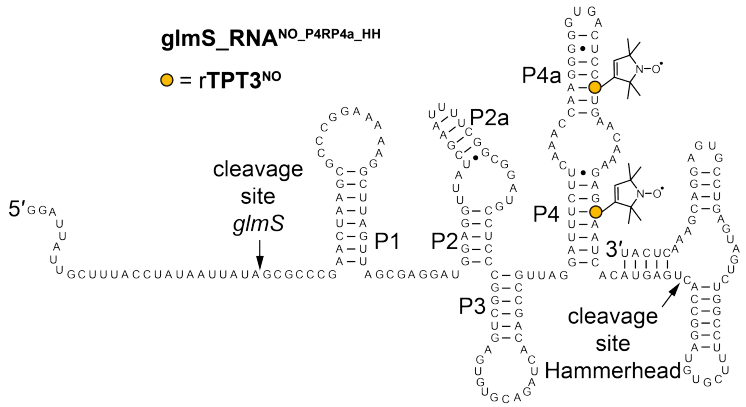


Figure 100. Schematic representation of the proposed **glmS_RNA^{NO_P4RP4a_HH}** ribozyme cassette sequence and secondary structure.

To achieve this goal, a four-piece PCR was envisioned in accordance to the former three-template approach (Figure 101A) as it is known to work in the same way for several overlapping DNA sequences[390]. A regular two-step assembly PCR (see section 5.3.4) employing canonical DNA templates was performed and yielded a full-length construct of desired 249 bp length as analyzed by agarose gel electrophoresis (Figure 101B). However, this was not the case for the reactions with an expanded genetic alphabet involving doubly d**NaM**-modified antisense template **glmS_DNA^{3_P4PR4a_NaM}**.

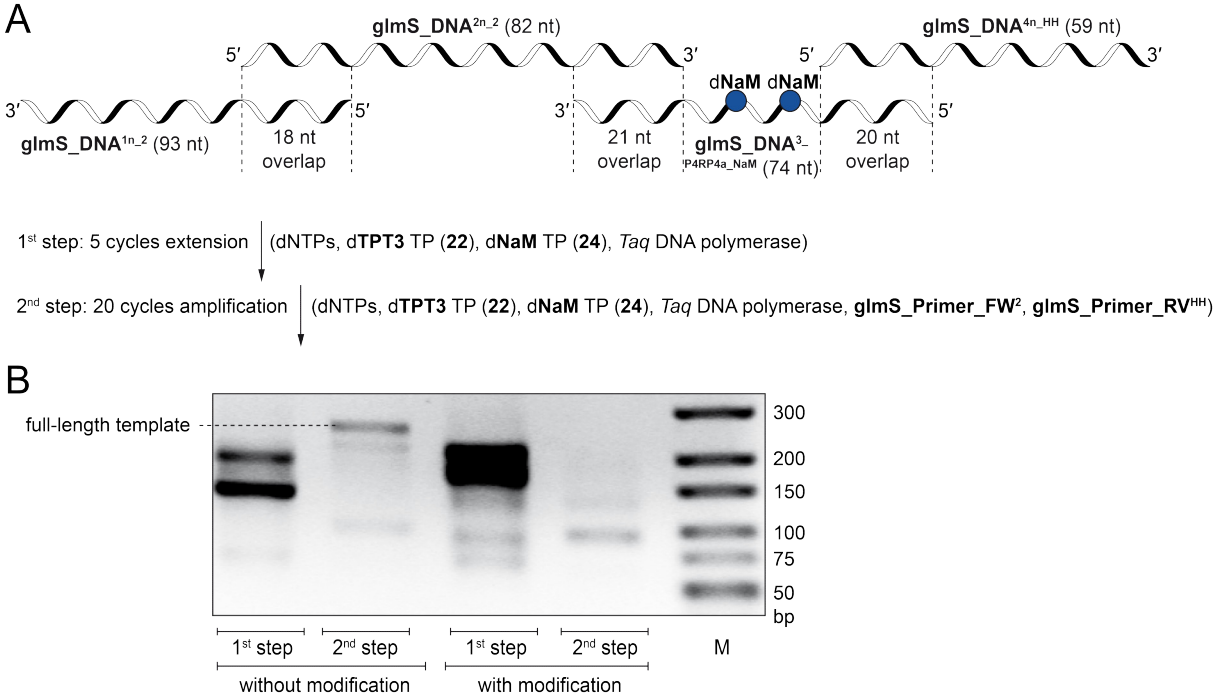


Figure 101. Attempted fusion PCR to generate full-length dsDNA templates for *in vitro* transcription of **glmS_RNA^{NO_P4RP4a_HH}** flanked by a Hammerhead ribozyme; A. Four overlapping templates were employed for the two-step assembly approach (shown is the UB-modified variant); B. Representative 2 % agarose gel (EtBr stain) of an unmodified and a UB-containing PCR; M = Marker.

An alternative two-step PCR program was applied[390] and optimized via gradient PCR. The new parameters basically implicated extended times for denaturation, hybridization, and elongation (see section 5.3.4). This should enable *Taq* DNA polymerase to successfully fulfill extension and amplification of the more complex oligonucleotide mixture and was further thought to be beneficial for (potentially slower) incorporation of unnatural deoxynucleotides.

Agarose gel analysis of the first step revealed extended sequences of considerable heterogeneity (Figure 102A, left panels), which is in agreement with the expectations as the polymerase produces several elongated template versions at this stage and is not able to produce the complete strands due to the lack of primers. Successful generation of clean full-length DNA using this method could then be shown after PCR amplifying the crude mixture in a second step mixture with appropriate primers added (Figure 102A, right panels).

These findings encouraged to perform a first test transcription. r**TPT3^{CP}** TP (**34**) was employed in an *in vitro* reaction from full-length dsDNA templates to facilitate sensitive readout of cyclopropene-modified RNA construct **glmS_RNA^{CP_P4RP4a_HH}** via IEDDA reaction with ^{Me}**Tet-BDP** (**59**).

Unspecific nucleic acid stain *SYBR Safe* on DPAGE separated samples revealed presence of the proposed ribozyme species and cleaved Hammerhead fragment (Figure 102B, left panels), however, no fluorescent signal could be detected (Figure 102B, right panels). Therefore, successful introduction of r**TPT3^{CP}** nucleotides could not be verified. Another band was observed closely below the commonly observed cleaved fragment of the *glmS* transcript, potentially representing a truncated version of the RNA. General feasibility to preparare *glmS*-Hammerhead RNA could be demonstrated by this method. Lack of fluorescence in the ribozyme core led to suspicion of UBP information loss during PCR amplification. At this point, insufficient 6-letter PCR could not be distinguished from unsatisfactory genetic alphabet expansion transcription. Moreover, decreased quality of one or several building blocks, *i.e.* d**TPT3** TP (**22**), d**NaM** TP (**24**), r**TPT3^{CP}** TP (**34**), and ^{Me}**Tet-BDP** (**59**), could be source of missing experimental success.

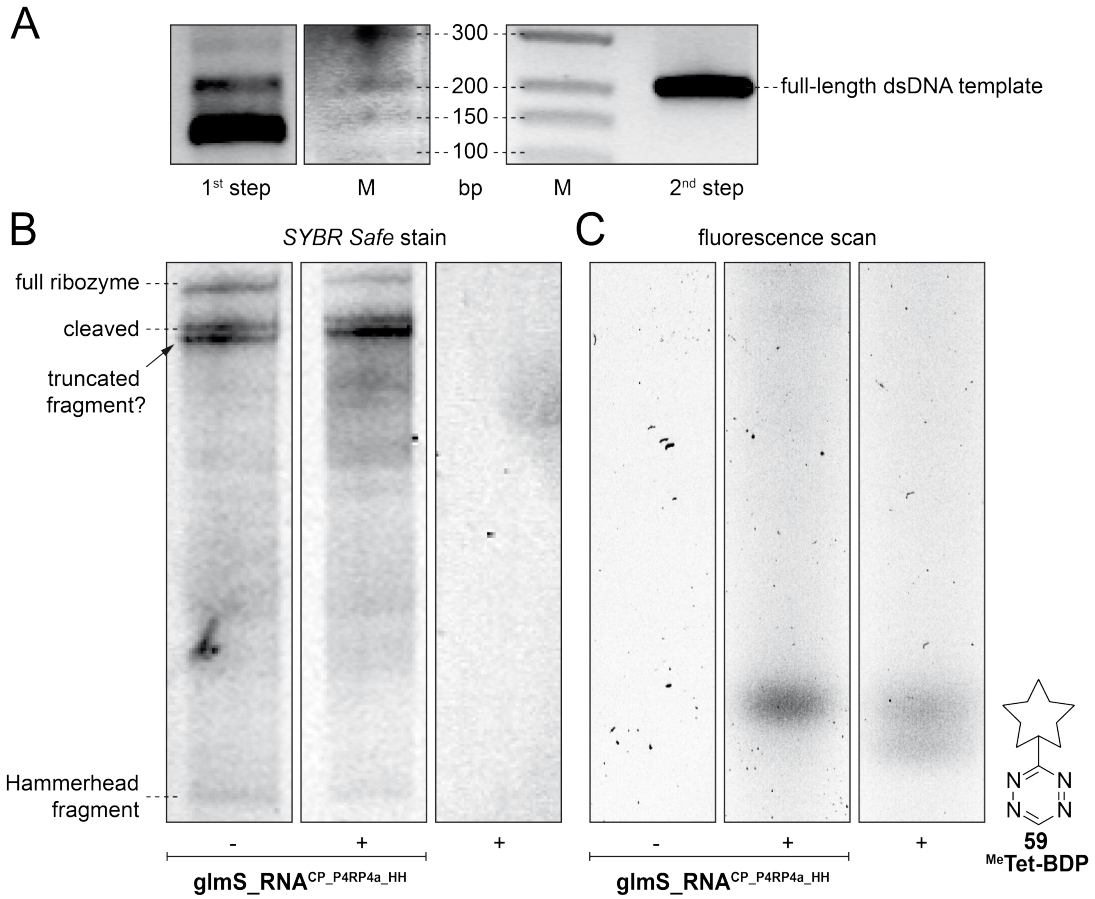


Figure 102. Gel electrophoretic analyses of DNA templates and transcripts to obtain **glmS_RNA^{CP_P4RP4a_HH}**; A. 2 % agarose gel (EtBr stain) of an UBP-modified two-step PCR with parameters adapted from Rydzanicz *et al.*[390]; B. 12 % DPAGE of T7 transcription reactions in the presence of r**TPT3^{CP}** TP (**34**) incubated with **^{Me}Tet-BDP** (**59** in *SYBR Safe* post staining (left panels) and fluorescence scan (right panels).

After preparation of a sufficient dsDNA template amount, transcription with r**TPT3^{NO}** TP (**61**) was performed. DPAGE separation from the cleaved Hammerhead fragment and electroelution of the desired RNA yielded 200 pmol **glmS_RNA^{NO_P4RP4a_HH}** for EPR analysis. Unexpectedly, this sample did not exhibit any nitroxide radical signal. Here, the DPAGE/electroelution purification process is suspected to be the predominant source of error as base loss of r**TPT3^{NO}** nucleobase might be a crucial factor (discussed in section 3.2.2.2). The presented approach intended exclusion of unspecifically attached r**TPT3^{NO}** nucleotides at the RNA's 3′-end. Therefore, reliable and quantitative removal of the cleaved Hammerhead sequence was required and could be reached via DPAGE separation in a straight-forward manner. Apparent stability problems of r**TPT3^{NO}** under these purification conditions unfortunately inhibited further investigations using this setup.

Unspecific 3′-extension of transcripts with r**TPT3^{NO}** TP (**61**)

To confirm or dissipate the suspicion that the high distance PELDOR peak of **glmS_RNA^{NO_P4RP4a}** originates from non-templated 3′-extension with r**TPT3^{NO}** nucleotides, reproduction of the sample was performed. Using 6-letter fusion PCR, subsequent expanded genetic alphabet transcription, and HPLC purification, a 1.0 nmol **glmS_RNA^{NO_P4RP4a}** sample could be obtained, which was analyzed via gel electrophoretic methods (Figure 103A/B).

EPR spectroscopic investigations of the spin labeled RNA resulted in similar data compared to the first attempt (see above). The *cw*-EPR spectrum (Figure 103C) again showed significant immobilization of the nitroxides, possibly in superposition with a more flexible species, and similar modulation was observed (Figure 103D). As in the first EPR experiment involving **glmS_RNA^{NO_P4RP4a}**, the PELDOR distance distribution exhibited a major peak at 4.5 nm (Figure 103E), which is in good accordance to the expected value. Relative to earlier measurements on this *glmS* construct, the major signal appeared sharper and included a shoulder at ~5–6 nm.

Using an optimized set of measurement parameters, the second peak at higher distances shifted from ~7 nm to higher values (8.0–8.5 nm) and showed decreased intensity with respect to the first attempt. These findings are in excellent agreement with the previously shown MD calculations. Together with an excellent spin labeling efficiency of ~88 % (c_{A260} = 34 µM theoretical nitroxide concentration versus c_{EPR} = 30 µM obtained by spin counting), the presented approach proved to be a powerful and viable technique studying intramolecular distances within complexly folded RNA molecules.

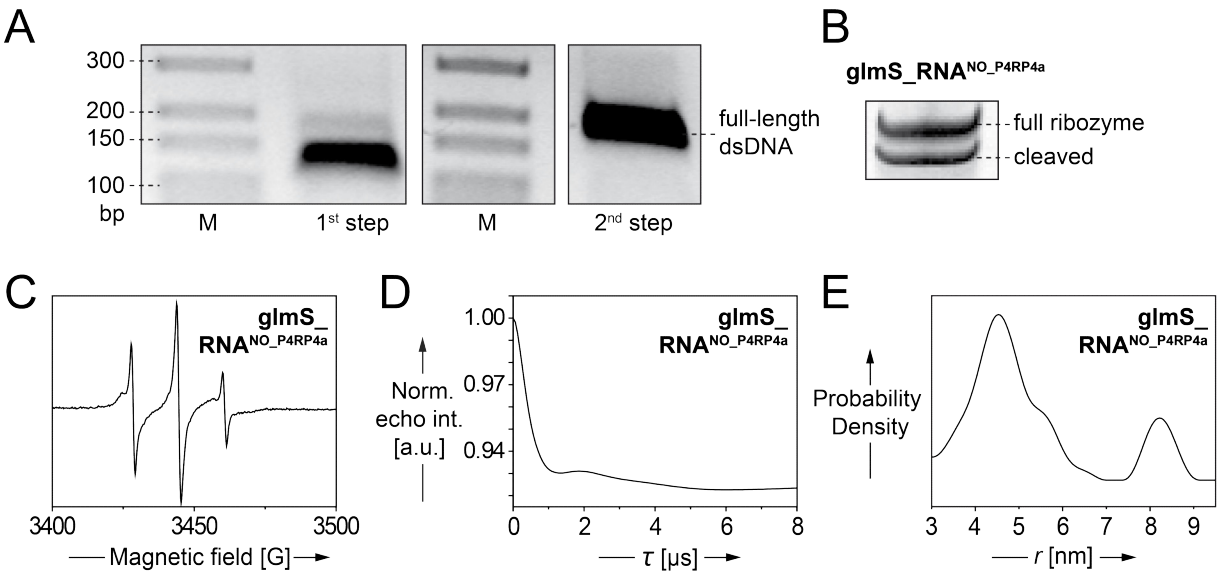


Figure 103. Successful reproduction of **glmS_RNA^{NO_P4RP4a}** EPR spectroscopic data shown in Figure 96; A. 2 % agarose gels (EtBr stain) of dsDNA template generation employing 6-letter fusion PCR; B. 12 % DPAGE analysis (*SYBR Safe* stain) of the **glmS_RNA^{NO_P4RP4a}** transcript; C. *cw*-EPR spectrum; D. Background-corrected PELDOR echo time trace (for raw data see section 5.3.18); E. PELDOR measurement with optimized parameters to accurately detect the high distance peak.

Since reproduction of **glmS_RNA^{NO_P4RP4a}** repeatedly gave a second, higher PELDOR distance value, this peak was verified to be truthfully correlated to the sample rather than representing a deconvolution artifact. Addressing 3′-incorporation of an r**TPT3^{NO}** nucleotide as the origin, another construct, **glmS_RNA^{NO_P4a}**, was prepared. This particular *glmS* variant was set out only to bear one of the former two spin label modifications. With a single r**TPT3^{NO}** entity introduced into stem P4a, observation of intramolecular inter-spin distances was not intended. In the case of significant non-templated 3′-extension with a spin labeled unnatural base, at least detection of the far distance of 7-9 nm was expected. The corresponding dsDNA template was assembled by fusion PCR (Figure 104A). T7 *in vitro* transcription yielded the desired construct **glmS_RNA^{NO_P4a}** as confirmed by DPAGE (Figure 104B). A 2.0 nmol HPLC-purified sample was submitted to EPR analysis and gave a spin labeling efficiency of ~90 % (c_{A260} = 20 µM, c_{EPR} = 18 µM). Indeed, distinct probability values were obtained by evaluation of a PELDOR echo time trace, which showed moderate modulation (Figure 104C and D). The most prominent peak appeared at 8-9 nm, which fits the P4a-3′ distance expectations. Figure 104D includes the distance distribution of **glmS_RNA^{NO_P4RP4a}** (dashed line). By the striking overlap of the high distance peaks, unspecific 3′-extension of *glmS* transcripts with r**TPT3^{NO}** entities was confirmed as origin for higher distance values. Interestingly, also further signals of decreased intensity are observed at ~4.5 nm and ~6.5 nm, which could arise from multiple stable conformations or intermolecular interactions between the ribozyme molecules.

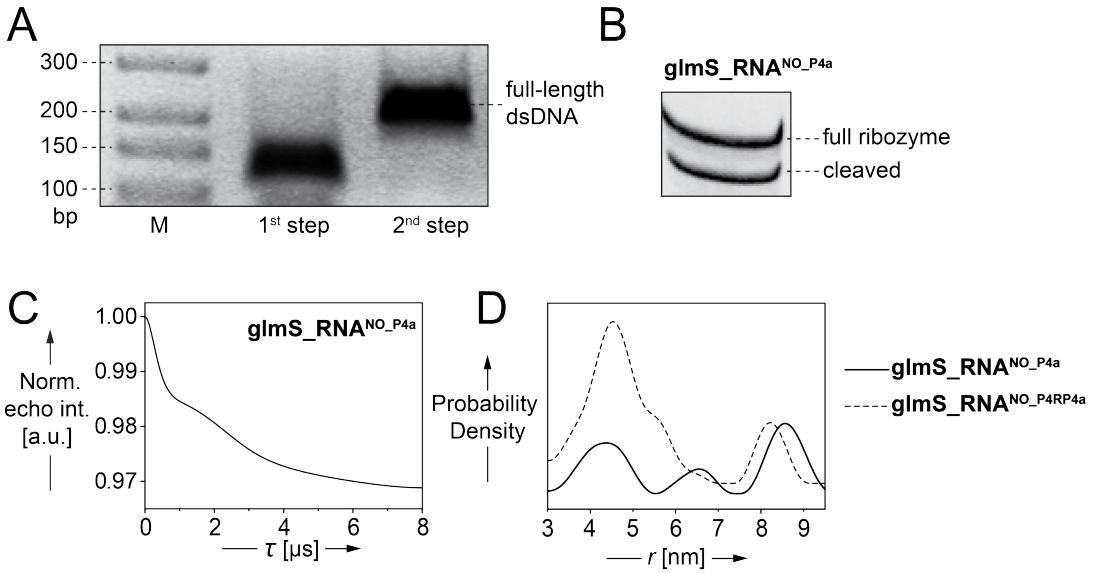


Figure 104. Analysis of **glmS_RNA^{NO_P4a}**; A. dsDNA template generation by two-step assembly PCR on a 2 % agarose gel (EtBr stain); B. Verification of the *glmS* variant RNA via 12 % DPAGE analysis (*SYBR Safe* stain); C. Background-corrected PELDOR echo time trace (for raw data see section 5.3.18); D. PELDOR distance distribution in overlay with the **glmS_RNA^{NO_P4a}** data shown in Figure 103.

For additional proof, the native sequence **glmS_RNA^{n1}** was transcribed from DNA containing only canonical nucleotides in an overnight reaction. r**TPT3^{NO}** TP (**61**) was added to the reaction mixture in the same concentration as applied in genetic alphabet expansion transcriptions. The obtained RNA was HPLC-purified to remove excess spin label triphosphate and a *cw*-EPR spectrum was recorded (Figure 105). An RNA concentration of c_{A260} = 25 µM was determined by UV absorption. Integration of the EPR spectrum gave an unexpectedly high spin concentration of c_{EPR} = 15 µM. Taking together these findings, unspecific incorporation events could affect 60 % of the transcribed RNA, which even exceeds the literature-known upper limit of 50 % non-templated 3′-extension[391].

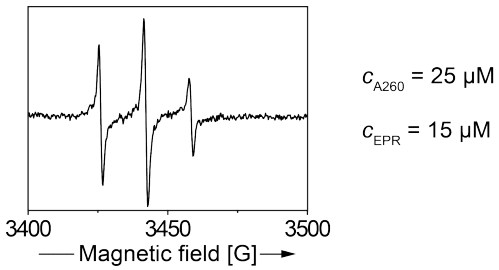


Figure 105. *cw*-EPR spectrum of **glmS_RNA^{n1}** transcribed in the presence of spin label triphosphate **61**.

The interestingly pronounced attachment of an r**TPT3^{NO}** nucleotide at the transcript end could have several reasons. Since transcription yields were generally decreased employing an expanded genetic alphabet, it could be assumed that unnatural **TPT3**:**NaM** nucleotides are processed considerably slower than canonical building blocks. Imperfect fitting of r**TPT3^{NO}** triphosphate **61** in the active site of T7 RNA polymerase may delay association and dissociation of the complex. This could result in prolonged accommodation time of **61** and would thus increase probability of unspecific incorporation events.

Implication of 2′-OMe-modified templates is suggested to be of essential interest studying large RNAs such as the *glmS* constructs with EPR. Efficient suppression of undesired 3′-incorporation in prospective experiments is expected by this technique as presented for short duplex transcripts in section 3.2.2.1. Furthermore, addressing the non-templated addition of r**TPT3^{NO}** nucleotides could involve a concentration dependency to investigate if the amount can be decreased by lowered **61** supply in the transcription reaction. Furthermore, correlation of unspecific 3′-extension and transcription time could possibly reveal a compromise between RNA yield and correct identity of desired sequences.

Future investigations on successful and reproducible distance measurements in highly structured RNA using genetic alphabet expansion transcription are currently performed in the Kath-Schorr group. In the following section valuable information acquired during this study to sustainably spin label target ribonucleic acid molecules by this novel technique are summarized with respect to future efforts.

3.2.2.4 Future perspectives of the project

The efforts of this study have proven site-specific spin labeling of RNA using r**TPT3^{NO}** TP (**61**) to be a versatile tool serving as nanometer range ruler under solution conditions. In order to gain reliable EPR data in a straight-forward manner, several aspects that crucially impact the approach were encountered during this work.

Design of target molecule's labeling positions should be performed with most possible care. Introduction of r**TPT3^{NO}** nucleotides in stem regions of highly structured RNA molecules will likely destabilize the helix when incorporated opposite a canonical base and could impair the integrity needed for proper distance measurements. Therefore, it should be beneficial to choose longer stems with high melting temperatures. Furthermore, introduction of the whole r**TPT3^{R}**:r**NaM** base pair into RNA should resolve this issue. However, this technique is not yet satisfyingly established and might suffer from inefficient r**NaM** incorporation by T7 RNA polymerase. Sections 3.4.2 and 3.4.3 will discuss this topic in deeper detail.

For short constructs, labeling efficiency and integrity of the modified RNAs is accessible via HPLC-MS. Unfortunately, this is not possible for transcripts exceeding ca. 50 nt. Positioning of each spin labeling position in the target ribonucleic acid sequence with approx. 35 nt canonical bases upstream in the 3′-direction would advantageously enable labeling quantification by means of a reverse transcription approach. This technique will be presented in section 3.4.3.

dsDNA templates involving the UBP **TPT3**:**NaM** for PCR amplification should best be addressable by standard sequencing methods, *i.e.* labeling positions should exhibit ≥80 nt up-or downstream. Hereby, hybridization with a proper 15-30 nt primer should be enabled and a safe primer-UB site distance should facilitate stably running commercial (*e.g.* Sanger) sequencing. The templates generated by six letter PCR could thus be analyzed with regard to UBP retention which is the fundamental prerequisite for efficient r**TPT3^{NO}** nucleotide incorporation during *in vitro* transcription. Moreover, d**NaM** nucleotides could be implicated in the reverse primers to ensure full modification in template DNAs. Such an approach is subject of current works. Recent literature suggests One*Taq* DNA polymerase to exhibit superior performance regarding **TPT3**:**NaM** propagation[332] compared to *Taq* DNA polymerase and will thus be implicated in future attempts.

To facilitate optimal transcriptional efficiency and fidelity, purity of r**TPT3^{NO}** TP (**61**) is essential. Further enhancement of unnatural nucleotide incorporation could be achieved using additives such as spermidine, Triton-X, or dimethylsulfoxide (DMSO). These substances are known to increase transcription performance of T7 RNA polymerase[48,335,396] and could thus contribute to expanded genetic alphabet applications involving **61**.

As 3′-unspecific incorporation of r**TPT3^{NO}** nucleotides seems to be a serious issue, suppression of this phenomenon is inevitable to gather reliable PELDOR distance data.

Design of DNA templates and target RNAs involving ribozyme cassettes that produce homogenous 3′-ends can serve as powerful tool achieving this goal.[48,89,366] Easy and straight-forward success could potentially be reached using 2′-methoxy (2′-OMe) modified reverse DNA or DNA primers at the 5′-end of antisense template strands.[48,368,369] Preliminary data of the Kath-Schorr group suggest outstanding performance applying such modifications. Thus future implication is strongly recommended for future RNA spin labeling experiments using r**TPT3^{NO}** TP (**61**).

The purification method of modified RNA samples is crucial for reliable experimental outcome employing EPR spectroscopy. Application of centrifugal membrane filtration (*Amicon Ultra*, *Merck Millipore*) was found best to ensure fast sample preparation, quantitative removal of excess triphosphates, and minimal impact on target sequences. As shown in section 3.2.2.2, preparative DPAGE is suggested to promote partial or quantitative loss of r**TPT3**-related nucleobases and is thus unsuitable.

Transcripts, which were truncated before r**TPT3^{NO}** incorporation, can be separated by HPLC to obtain homogenous sample material. This is, however, only possible for relatively short fragments and applicability to longer RNAs might be infeasible. Contamination with such unlabeled RNA will impact concentration termination as well as hybridization events. Thus, consideration of this issue is required to achieve effective experimental design.

Since occurrence of r**TPT3^{NO}** nucleotide decomposition (section 3.2.2.2) is not yet deeply understood, the choice of HPLC buffer is also suggested to be of central importance. It is assumed, that evaporation of samples might have an acidifying effect employing *e.g.* TEAA buffer. Bicarbonate-based alternatives should be employed in future efforts avoiding the risk of base loss.

Embracing these findings and suggestions in ongoing research will likely improve r**TPT3^{NO}** spin labeling techniques to gain structural information on complex RNA molecules employing EPR spectroscopic methods in a short-term manner.

3.3 Towards a photoswitchable unnatural base

Aiming on direct introduction of functional groups via genetic alphabet expansion transcription, creation of an unnatural base bearing a photoswitchable unit was investigated during this study. Such a compound is supposed to enable site-specific introduction of corresponding ribose building blocks into oligonucleotides. Thus, the intriguing opportunity to spacio-temporally control various RNA interactions in a reversible manner would be given. It is expected that light-triggered ring closure and reopening of a diarylethene scaffold (as presented in section 1.5.3) attached to RNA would substantially alter interaction surfaces and consequently promote or inhibit nucleic acid performance.

Conveniently, the **TPT3**-related nucleobase analog **TPT2** by the Romesberg group[194] (Figure 106A) is constituted of a pyridinethione-fused methylthiophene and thus provides the necessary structural basis for the construction of a diarylethene-derived photoswitchable nucleoside (Figure 106B/C, highlighted in red and blue).

In the following sections different synthetic pathways to successfully establish facile preparation of an UV-switchable unnatural nucleoside will be reported and discussed.

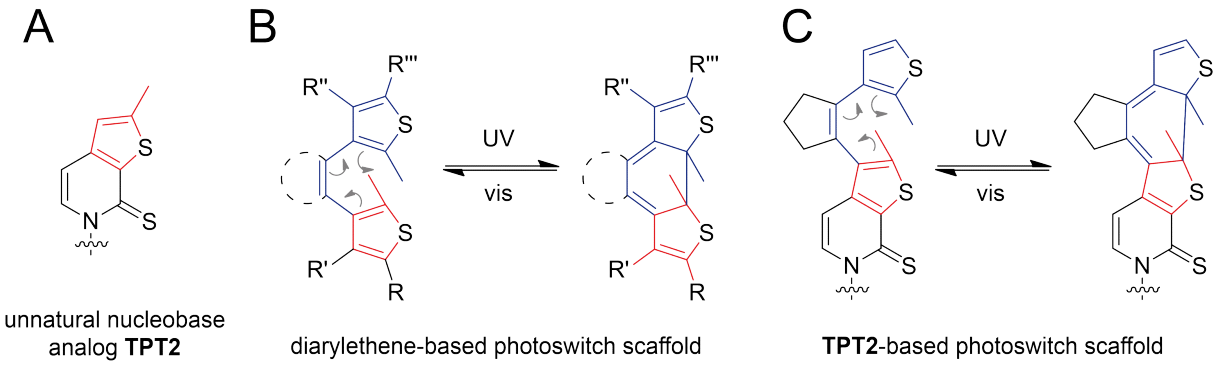


Figure 106. Fundamental considerations to develop a photoswitchable unnatural base scaffold; A. Romesberg's unnatural **TPT2** nucleobase residue[194] with methylthiophene entity (red) as building block for build-up of a photoswitch; B. Schematic representation of diarylethene-based photoswitches (red/blue); C. A photoswitchable unnatural nucleobase moiety based on the **TPT2** unnatural nucleobase; sugar moieties are omitted for clarity.

3.3.1 Synthetic access to a novel photoswitchable nucleobase analog scaffold

A **TPT2**-based photoswitchable unnatural base ribonucleoside, r**TPT2**^{PS} (**62**), was aspired. **62** was retro synthetically disassembled into peracylated β-D-ribose **37** and dithienylcyclopentenyl pyridone **63** (Figure 107). Preparation of the cyclopentene unit of **63** was planned from 1,5-pentanedione derivative **64** via an intramolecular McMurry reaction. Further segmentation revealed three building blocks attachable to each other via Friedel-

Crafts acylation: thienopyridone **65** as **TPT2** nucleobase analog precursor, glutaric acid (**66**), and 2-methylthiophene (**67**).

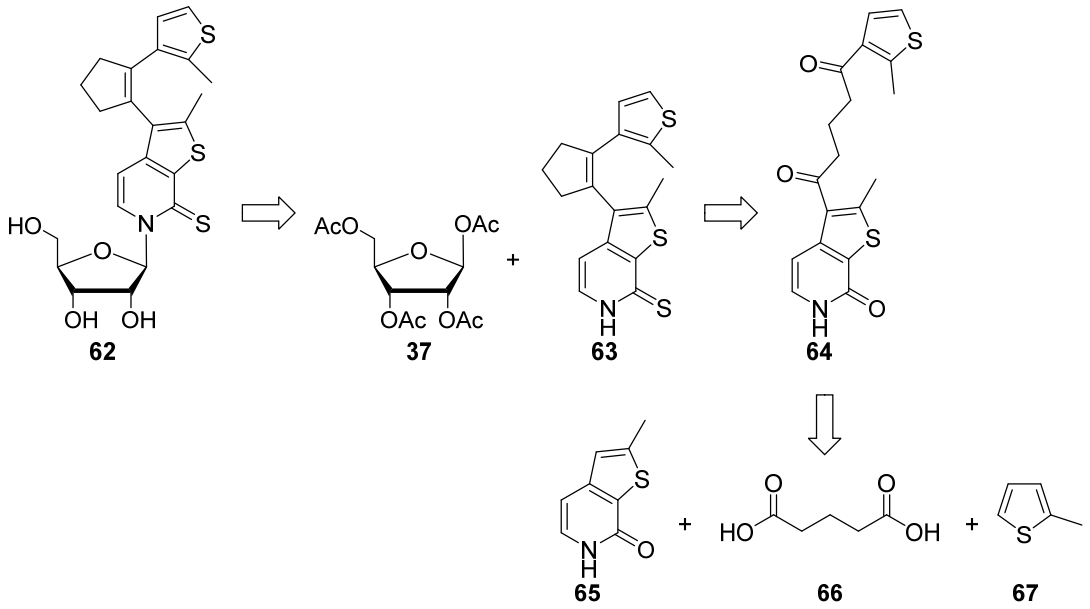


Figure 107. Retro synthetic considerations to develop photoswitchable unnatural base nucleoside r**TPT2^{PS}** (**62**) via preparation of the nucleobase analog scaffold **63** by a Friedel-Crafts-based approach.

3.3.2 A Friedel-Crafts-based strategy

In order to enable synthesis of the proposed photoswitchable nucleoside r**TPT2^{PS}** (**62**), decoration of methylated thienopyridone scaffold **65** with a suitable electron withdrawing group (EWG) under basic conditions was intended. Thus, correct reactivity towards substitution and protection of the amide function in subsequent reaction steps were expected (Figure 108). Friedel-Crafts acylation with glutaric anhydride (**68**) leading to a pentanoic acid derivative as aspired. *In situ* conversion of the carboxylic acid into the acid chloride was assumed to facilitate a further Friedel-Crafts acylation allowing attachment of a thiophene compound. To avoid drawbacks in yield arising from unselective acylation, application of dimethylated thiophene **69** was favored over initially proposed 2-methylthiophene (**67**). Being aware that the final product will slightly differ from the initially proposed diaryl cyclopentene scaffold, negative impact on the photoswitching properties was not expected.

The 1,5-diketone entity constructed by Friedel-Crafts reactions was intended to yield a 1,2-substituted diaryl cyclopentene under McMurry conditions, developing the photoswitchable scaffold. Removal of the introduced protecting group in basic aqueous medium was planned to eventually give unnatural diaryl cyclopentene nucleobase analog precursor **63**.

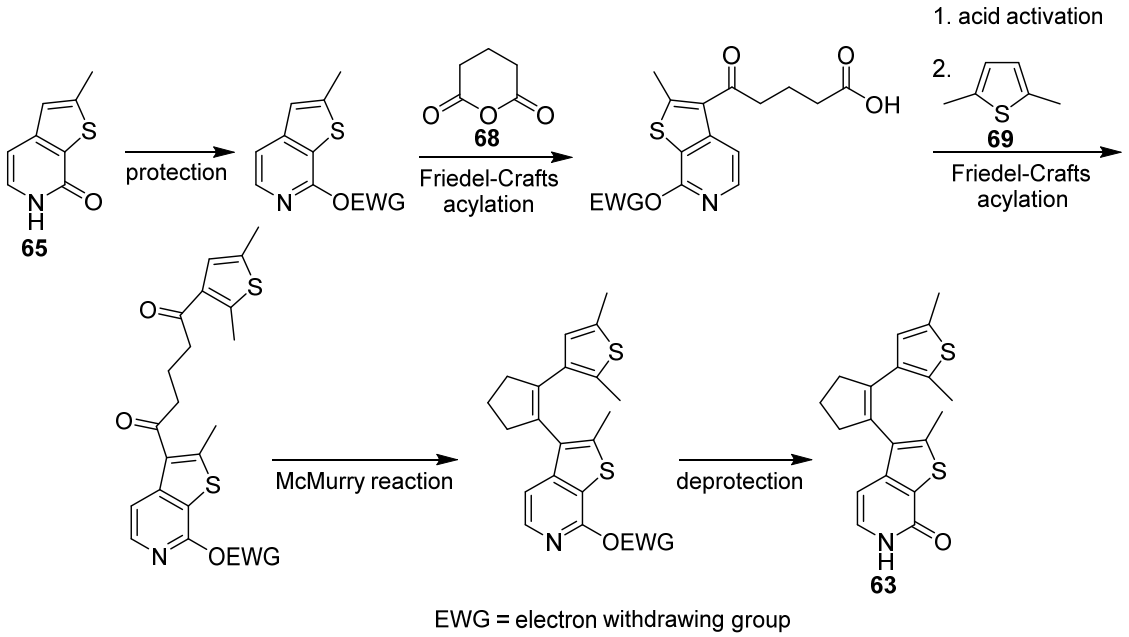


Figure 108. Proposed synthetic scheme leading to the novel nucleobase analog precursor **63** within six steps starting from methylated pyridone derivative **65**.

Synthesis of thienopyridone derivative **65**[194]

The synthesis of **TPT2** nucleobase analog precursor **65** was based on a procedure by Li *et al*.[194] and follows similar steps as described for thienopyridone **20** (section 3.1.1.2), starting from 5-methylthiophene-3-carboxaldehyde (**70**).

Commercial supply of **70** is provided at ~230 \$ g^{-1} (source: *Scifinder*, accessed 2019/04), thiophene-3-carboxaldehyde (**18**) in contrast is available from ~2 \$ g^{-1} (source: *Scifinder*, accessed 2019/04). This compound is convertible into its 5-methyl derivative by a three-step one-pot synthesis[397] (Figure 109); therefore, in lab preparation of **70** was favored.

Methylation of **18** can be achieved via a reaction cascade involving α-amino lithoxide formation using *N*-methylpiperazine and *n*-butyllithium (*n*-BuLi) followed by *N*,*N*',*N*'',*N*'''-tetramethylethylenediamine (TMEDA)-mediated lithiation with *sec*-butyllithium (*s*-BuLi) and subsequent alkylation by methyl iodide (MeI). The reaction provides a satisfactory regioselectivity of 8:2 towards the desired 5-methyl isomer.

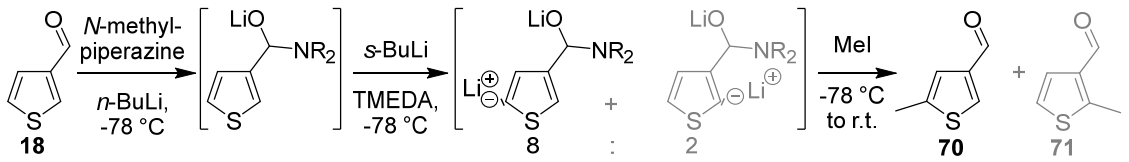


Figure 109. Schematic methylation mechanism of **18** providing semi regioselective access to **70**.

Synthesis of 5-methylthiophene-3-carboxaldeyde **70** resulted in a mixture with 2-methylthiophene-3-carboxaldehyde (**71**) and the starting material thiophene-3-carboxaldehyde (**18**) after aqueous workup. A **70**:**71**:**18** ratio of 65:17:18 was achieved according to integrals of their aldehyde protons in ^{1}H-NMR spectroscopic analysis. 82 % of total material could be isolated, containing approx. 53 % product **70**. Separation by silica gel chromatography or distillation both left the shares of the compounds unimproved. It was hypothesized, that further synthetic steps would allow for removal of side products and synthesis was continued as scheduled (Figure 110)

Knoevenagel condensation with malonic acid (**44**) resulted in full conversion of the isomeric mixture to acrylic acids **72** and **73**.

In situ conversion into acid azides facilitated Curtius rearrangement and intramolecular cyclization at elevated temperatures to give 45 % of **65**. 2-methylated thiophene acrylic acid **73** cannot undergo electrophilic substitution at the thiophene α-position and is removed during workup. Minor impurities in the final 2-methyl substituted thienopyridone potentially arise from the presence of unmethylated species that was carried along in the reaction cascade and do not interfere with following synthetic procedures.

65 could be obtained in 24 % overall yield in seven steps from thiophene-3-carboxaldehyde (**18**) and the presented approach is a convenient way to affordably provide the product in gram scale.

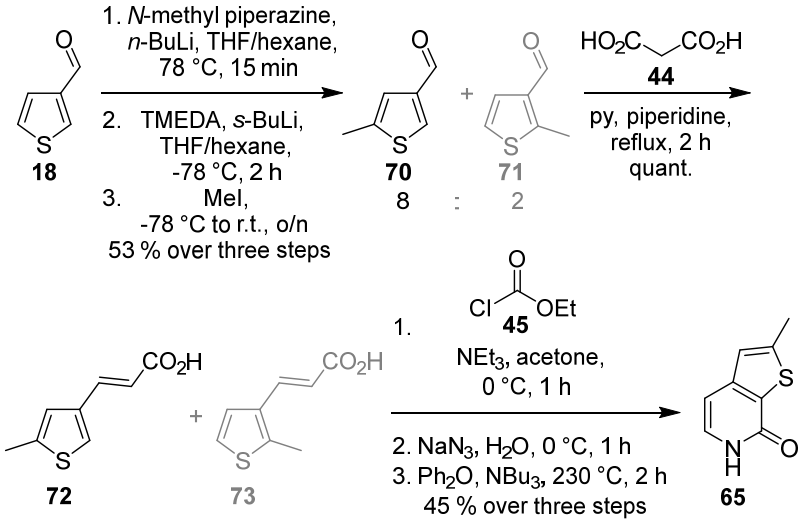


Figure 110. Synthesis of nucleobase analog **TPT2** precursor **65**.

Following, the introduction of a suitable protecting group was investigated. Reactivity of thienopyridone **65** towards electrophilic substitution is strongly directed in *meta* position relative to the amide nitrogen atom as it was performed for the iodination reaction during the synthetic route of r**TPT3^{CP}** TP (**34**) (see 3.1.1.3). To circumvent this inconvenience intermediate protection with electron withdrawing groups such as triflate, mesylate, or tosylate was envisaged in order to favor substitution at the thiophene moiety (Figure 111).

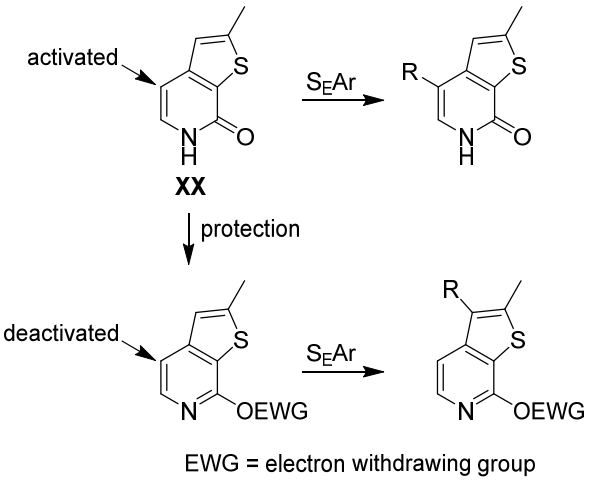


Figure 111. Postulated reaction cascade to enable regioselective electrophilic aromatic substitution (S_EAr) at the thiophene scaffold of **65**.

Several attempts to introduce electron withdrawing moieties were made (Figure 112). Upon addition of a 4-fold excess trifluoroacetic acid anhydride (TFA_2O) to **65** at 0 °C the initially turbid suspension directly became clear indicating a fast reaction. However, following the reaction by TLC, conversion to **74** could not be verified. After four days reaction time, chromatographic purification was applied and only starting material could be recovered. This suggests low reactivity towards TFA_2O and/or reversion of the reaction during workup (Figure 112a).

Employing an excess (approx. 3-fold) *N*-Phenyl-bis(trifuoromethanesulfonimide) to **65** at room temperature to give trifluoromethanesulfonyl(triflate/Tf)-modified **75** (Figure 112b) also did not result in significant conversion even after several days according to TLC.

Triflate derivatization could eventually be achieved by reaction with TfCl (10 eq.) at 0 °C and **75** was obtained with a 76 % yield after workup and purification (Figure 112c).

Alternatively, a methanesulfonyl (Ms) group could be introduced using MsCl (2.2 eq.) at 0 °C which enabled quantitative isolation of **76** (Figure 112d).

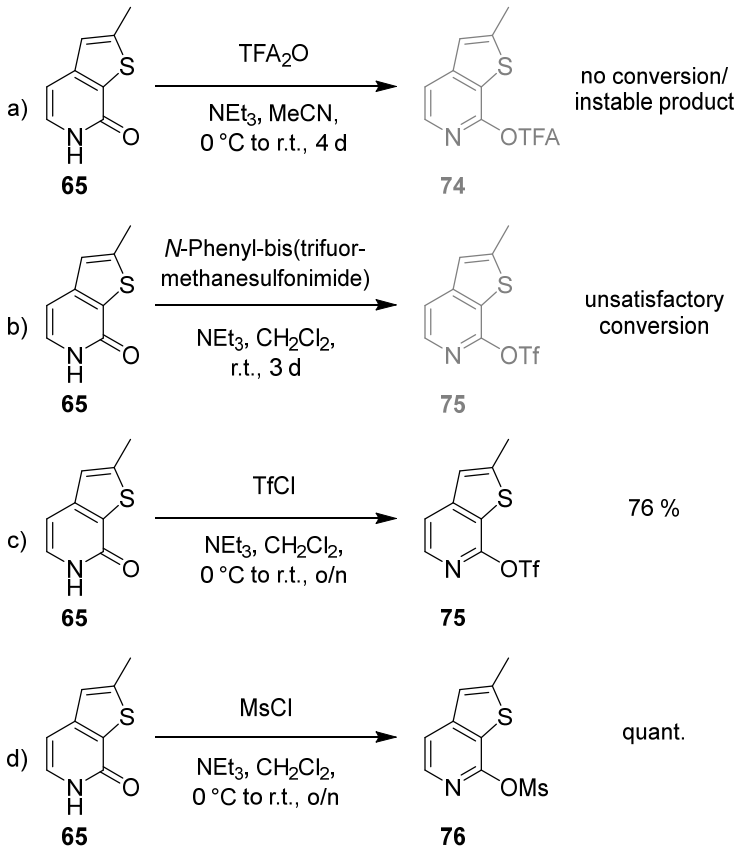


Figure 112. Introduction of protective and directing groups to **65**.

Unfortunately, Friedel-Crafts-type acylation of triflate-protected thienopyridone **75** with glutaric anhydride (**68**) could not be achieved (Figure 113). Despite a visible change from a pale beige to dark reddish/brownish solution after addition of Lewis acid catalyst $AlCl_3$, no conversion could be verified. Aqueous workup and chromatographic purification of the crude reaction led to recovery of starting material **75**.

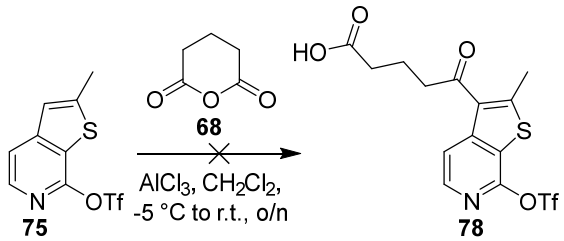


Figure 113. Unsuccessful attempt to synthesize 5-ketopentanoic acid derivative **78** by reaction of **75** with glutaric anhydride (**68**) under Friedel-Crafts conditions.

In another attempt, one carboxylic acid function of glutaric acid (**66**) was chlorinated *in situ* to give **77** and subsequently subjected to Friedel-Crafts conditions with **75** (Figure 114). However, NMR analysis of the crude mixture after overnight incubation revealed an unsuccessful transformation.

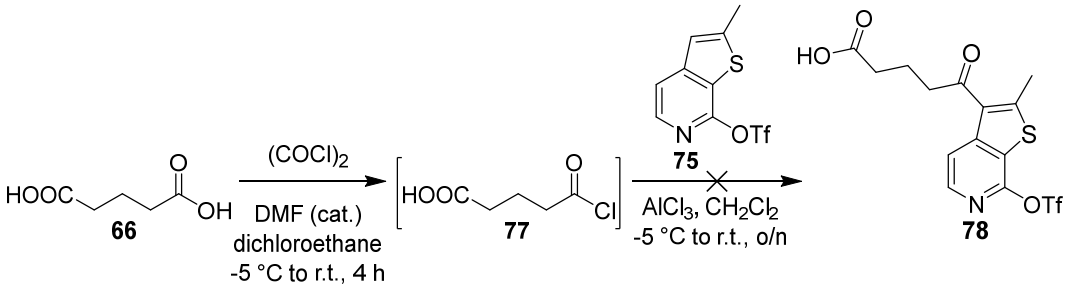


Figure 114. Inaccessibility of **78** via *in situ*-generated glutaric acid chloride (**77**) subjected to react with Tf-derivatized thienopyridone **75** under Friedel-Crafts conditions.

Next, the Friedel-Crafts acylation of 2,5-dimethylthiophene (**69**) with glutaric anhydride (**68**) was performed (Figure 115) to give 5-ketopentanoic acid derivative **79** in good yield (83 %) under standard conditions. This verifies the general feasibility and applicability of such reactions employing the parameters stated.

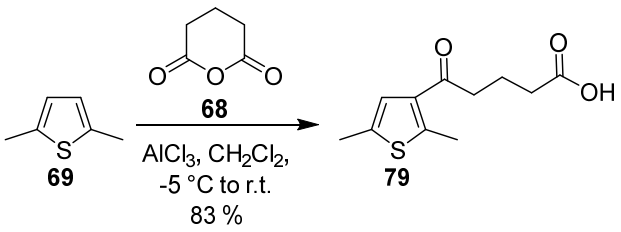


Figure 115. Friedel-Crafts acylation of 2,5-dimethylthiophene (**69**) with glutaric anhydride (**68**).

Triflate-protected thienopyridone derivative **75** appeared to be non-reactive in the presence of activated carboxylic acids and the powerful Lewis acid $AlCl_3$; therefore, mesylated species **76** was submitted to Friedel-Crafts conditions with the *in situ* generated acid chloride **80** (Figure 116). After overnight incubation, aqueous workup and silica gel chromatography only the deprotected **TPT2** nucleobase analog precursor **65** could be isolated (in a quantitative manner).

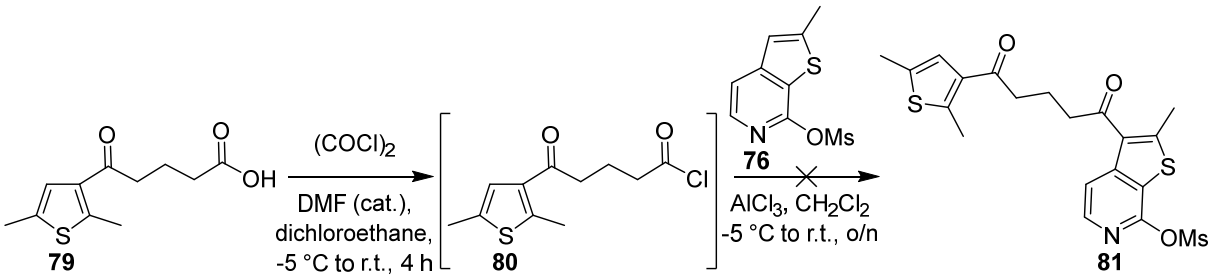


Figure 116. Attempted Friedel-Crafts coupling of *in situ* prepared acid chloride **80** and mesylated thienopyridone **76** to give the fused diketone **81**.

Under all tested reaction conditions, no formation of desired Friedel-Crafts acylation products involving protected derivatives of thienopyridone **65** could be proved. It is assumed, that the reactivities of the protective group-modified compounds **75** and **76** are insufficient for successful electrophilic aromatic substitutions.

Further experiments could involve investigation of reactions at elevated temperatures. Additionally, testing of alternative (Lewis) acid catalysts such as trifluoromethanesulfonic acid (TFMSA), $InCl_3$, or boron triluoride etherate, $BF_3{\cdot}OEt_2$, could possibly lead to effective conversion. Generating a 1,5-diketone scaffold such as **81** would open the possibility to elegantly prepare **TPT2** unnatural base analogs with diarylcyclopentene photoswitching properties by intramolecular McMurry reaction.

As an alternative, build-up of the nucleoside analog scaffold with transition metal-catalyzed cross-coupling reactions as key steps is very promising and will be discussed in the following section.

3.3.3 Cross-coupling-based approaches

Since derivatization of **TPT2** nucleobase analog precursor **65** remained unsuccessful, another retrosynthetic pathway was developed (Figure 117). Photoswitchable compound **63** could possibly be assembled in a transition metal-catalyzed manner from brominated thienopyridone **82** as well as suitably substituted cyclopentene and methylthiophene scaffolds. Here, cross-coupling reactions such as Heck reactions or Suzuki couplings are particularly promising transformations.

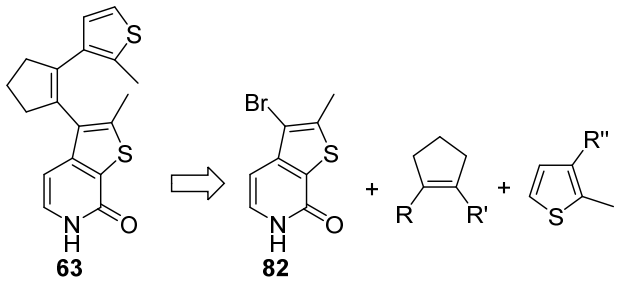


Figure 117. Retrosynthetic considerations to develop photoswitchable unnatural base precursor **63** via a transition metal-catalyzed approach involving brominated thienopyridone **82** as well as suitably substituted cyclopentene and 2-methylthiophene building blocks.

3.3.3.1 Towards creation of a novel unnatural nucleoside precursor

Synthesis of the brominated **TPT2** nucleobase analog precursor **82** can enable a variety of downstream functionalization possibilities.

Straight-forward preparation of compound **82** was aspired starting from 5-methylthiophene-3-carboxaldehyde (**70**) in a five-step approach (Figure 118). Bromination with *N*-bromosuccinimide (NBS) was intended to build up the necessary thiophene scaffold **83**. In analogy to the synthesis of thienopyridone **20** (see 3.1.1.2), *in situ* conversion into the corresponding acrylic acid **84** with a subsequent three-step cascade was planned eventually giving access to the cyclized product **82**.

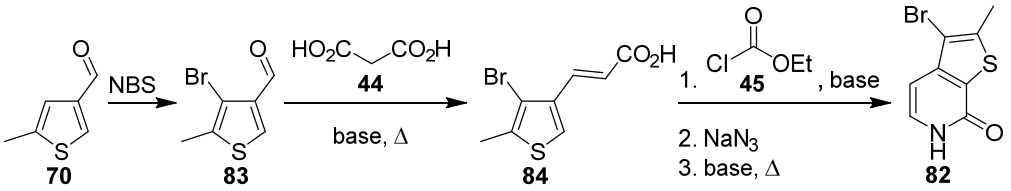


Figure 118. Proposed synthetic scheme towards brominated nucleobase precursor **82**.

5-Methylthiophene-3-carboxaldehyde **70** was prepared as stated above (section 3.3.2) and obtained as a mixture with 2-methyl derivative **71**. The subsequent synthetic procedure allows separation from undesired isomers; therefore, synthesis was continued as scheduled (Figure 119).

Bromination of **70** with NBS in an overnight reaction gave 14 % of 4-bromo-5-methyl-thiophene-3-carboxaldehyde (**83**) after purification. The low yield is mainly due to the various byproducts originating from the mixture of starting materials as well as a lack of regioselectivity of the reaction. It was, however, possible to successfully separate the desired product **83** from the complex crude product.

Acrylic acid **84** could then be isolated with usual good yield of 90 % by standard Knoevenagel condensation with malonic acid (**44**).

Intriguingly, a three-step approach involving intermediate generation of a mixed anhydride with ethyl chloroformate (**45**), conversion into the corresponding acid azide **85** and thermally induced Curtius rearrangement followed by intramolecular cyclization was unsuccessful. Neat acid azide **85** (78 % starting from **84**) was isolated in subsequent attempts ensuring purity and integrity of starting material. Various attempts under optimal reaction conditions (Ph_2O, NBu_3, 230 °C) did not result in any product formation and the presence of brominated thienopyridone **82** could not be verified, neither by NMR nor HPLC-MS analysis. Since the cyclization event relies on an ionic mechanism, reactivity could potentially be increased by choosing an apolar solvent. However, subjecting **85** to NBu_3 in refluxing toluene did not lead to the desired product. Pokhodylo *et al.* described formation of thienopyridones from acrylic

acid azide-substituted thiophenes in 1,2-dichlorobenzene at elevated temperatures adding catalytic amounts of iodine.[308] Submission of **85** to 1,2-dichlorobenzene at 140 °C for 2 h and subsequent refluxing at 180 °C for 2 h under the addition of iodine did not reveal any cyclization products.

Eventually, synthetic attempts towards **82** remained inconclusive. Presumably, the bromine substituent on the thiophene ring is vastly altering its chemical properties, inhibiting the intramolecular cyclization reaction.

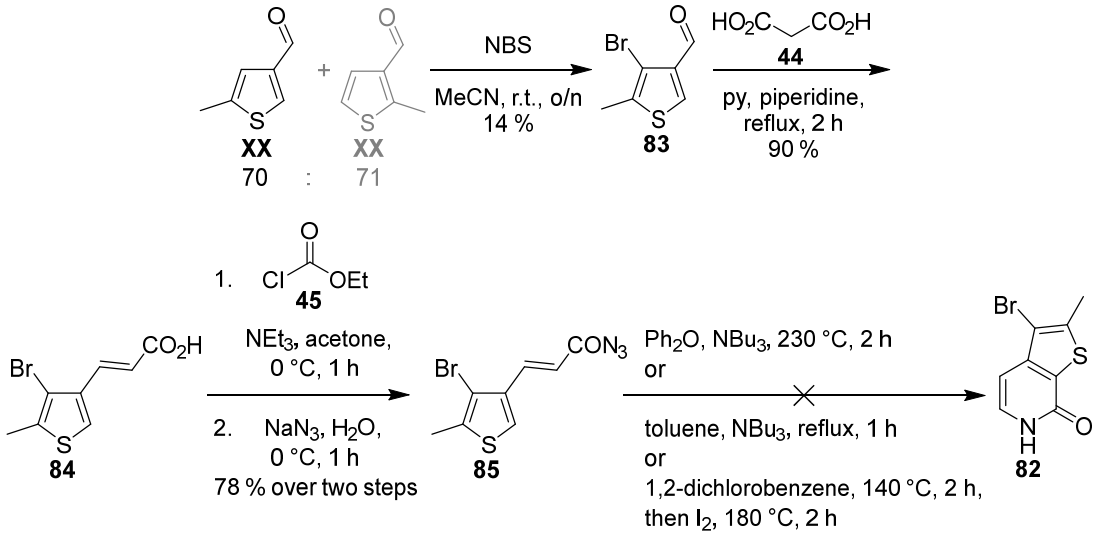


Figure 119. Attempted synthesis of brominated nucleobase analog precursor **82** via preparation of acid azide **85**.

Hornberger *et al.* described a microwave-assisted approach successfully brominating a similar compound, **86**[398], which resembles an unmethylated thiophenyl isomer of triflate-derivatized **75** (Figure 120), by irradiation in AcOH at 90 °C yielding 55 % of brominated **87**.

This suggests that triflate derivatization of thienopyridones in fact enables selective thiophene substitution. The harsh reaction conditions described in the literature further state low reactivity of compound **86** (and likewise of **75**) to undergo electrophilic aromatic substitution at the thiophene moiety. In the absence of a microwave reactor, further studies regarding the synthesis of **82** were inhibited.

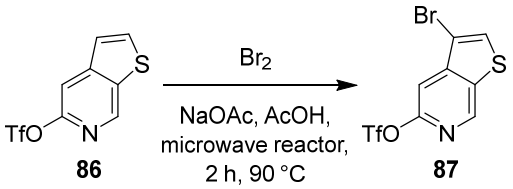


Figure 120. Microwave-assisted bromination of **86** by Hornberger *et al.* (literature yield: 55 %).[398]

Eventually, commercial supply of thienopyridone **82** could be obtained from *Chemspace* (Riga, Latvia). Final costs of ~1500 € g^{-1} strongly support a considerably challenging synthetic route. Nevertheless, this purchase enabled reception of **82** in a significant amount that allowed for further progress towards a photoswitchable nucleoside analog.
Prospective studies should definitively involve evaluation of possibilities for cost-effective in-laboratory preparation of thienopyridone **82** maintaining future works on this project.

3.3.3.2 Studies on the synthesis of a photoswitchable unnatural base nucleoside using Heck-type reactions as key steps

In order to synthesize a suitable building block attachable to the brominated unnatural nucleobase precursor **82**, a synthetic route based on Heck reactions was investigated. Fusion of **82** with a thiophenyl-substituted cyclopentene would allow generation of a photoswitchable nucleobase analog entity.
Heck reactions are usually conducted using aryl halides and acyclic olefins as coupling partners. A Pd(0) species acts as catalyst provided with phosphine ligands.[399,400] Great progress has been achieved in the past decades making this reaction applicable to cyclic alkenes in a regioselectively controlled manner.[401,402] With this technique, facile and affordable build-up of photoswitchable nucleobase analog precursor **63** was envisioned (Figure 121). Preparation of 3-bromo-2-methylthiophene (**88**) was intended via a *di*bromination and *de*bromination cascade starting from methylthiophene **67**. A first cyclic Heck reaction with cyclopentene (**89**) was planned, giving the methylthiophenyl-cyclopentene scaffold **90**. Subsequently, a second Heck-type approach was likely to fuse **90** and the brominated **TPT2** nucleobase analog precursor **82**, giving the potentially photoswitchable compound **63**.

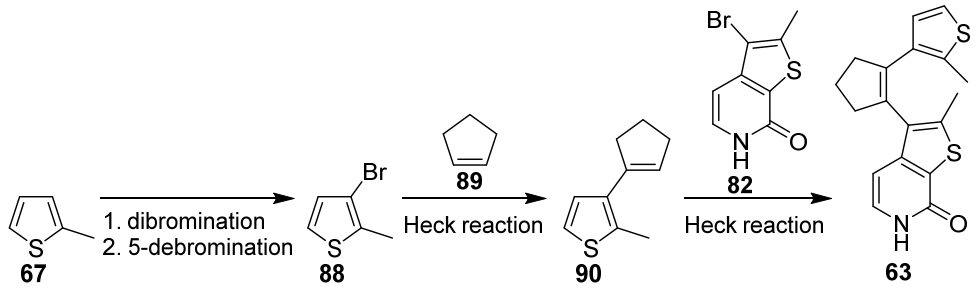


Figure 121. Proposed synthetic route towards **63** based on Heck reactions with cyclic olefins.

The promising straight-forward synthetic route was initiated with an intermediate dibromination of 2-methylthiophene (**67**) using NBS in acetic acid in a simple overnight reaction to give 88 % of 3,5-dibromo-2-methylthiophene (**91**) (Figure 122). Selective lithiation of the 5-position followed by aqueous workup quantitatively led to 3-bromo-2-

methylthiophene (**88**). In contrast to the Friedel-Crafts approach presented in section 3.3.2 this setup facilitates the use of singly methylated thiophene instead of the bulkier 2,5-dimethylthiophene. Analogous to literature procedures[402], a Heck reaction with cyclopentene (**89**) in DMF was performed. Palladium(II) hexafluoroacetylacetonate was employed as catalyst and tri-*tert*-butylphosphonium tetrafluoroborate served as ligand, DIPEA was added as non-nucleophilic nitrogen base. Heating the reaction to 120 °C for 2 h enabled isolation of the desired cyclopentenyl methylthiophene **90** (22 %).

An overall yield of 19 % was achieved. No attempts were made to optimize the reactions and the amounts obtained were sufficient to investigate applicability of downstream reactions. As simplest improvement prolonged reaction time of the Heck reaction should be considered.

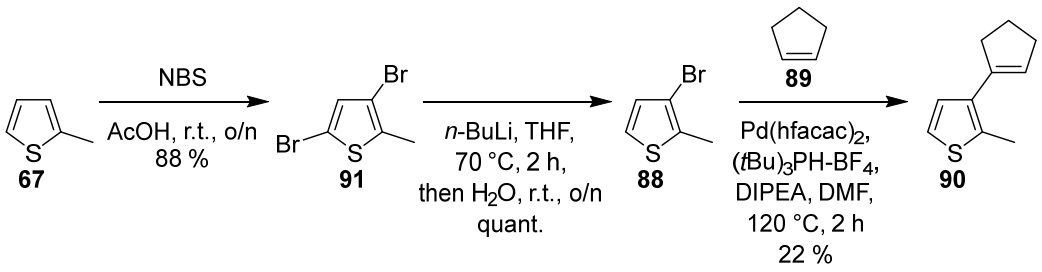


Figure 122. Synthesis of cyclopentenyl methylthiophene **90**.

Attempts to join **90** and the brominated **TPT2** nucleobase analog precursor **82** under similar reaction conditions did not succeed (Figure 123A). Applying chromatographic purification on the crude mixture, the only reasonable fraction that was obtained revealed debromination of **82** by ^{1}H-NMR analysis, giving the **TPT2** unnatural base precursor **65** (Figure 123B). A quartet appearing at 7.05 ppm corresponds to the thiophene proton. In combination with a doublet signal of the thiophene methyl group at 2.63 ppm, both exhibiting a coupling constant of 1.1 Hz, loss of the bromine substituent is clearly confirmed. This might indicate successful oxidative addition of the active Pd(0) species during the reaction, however subsequent mechanistic steps are inhibited.

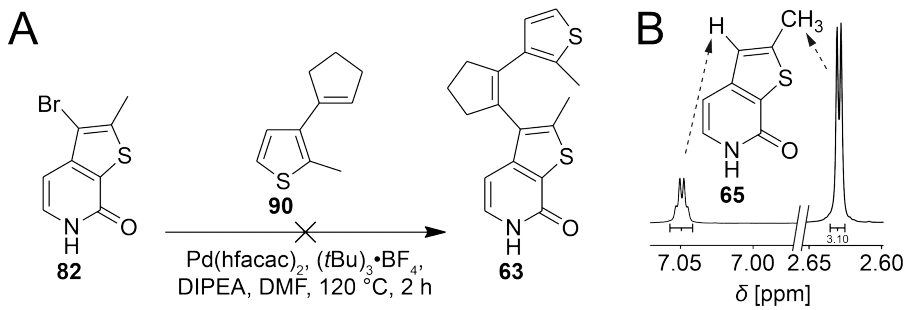


Figure 123. Attempted Heck reaction to fuse brominated **TPT2** unnatural nucleobase precursor **82** and cyclopentenyl methylthiophene **90**; A. Reaction scheme; B. Sections of ^{1}H-NMR (CD_3OD, 400 MHz) analysis of the chromatographically separated reaction mixture with indicative signals of **65**.

To test general accessibility of **82** to Heck reactions, a fusion between the thienopyridone and cyclopentene (**89**) was attempted (Figure 124). Product formation could not be observed, even when adjusting the conditions to an overnight reaction (~16 h) at 120 °C. Chromatographic separation of the reaction mixture enabled recovery of 78 % starting material **82**. Interestingly, loss of the bromine substituent did not seem to occur in this case.

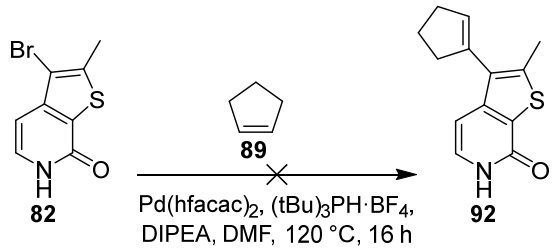


Figure 124. Submission of **82** and cyclopentene (**89**) to Heck conditions in an overnight reaction did not result in the desired diarylcyclopentene **92**.

The success of Heck reactions sensitively depends on electronic and steric effects[79, 80] which are often not fully predictable. Furthermore, a plethora of different mono- and bidentate phosphine ligands exists[403,404] as well as the choice of ligand might be essential for effective conversion. Also solvents, solvent mixtures and the base employed are severe factors controlling the reaction outcome.

Avoiding laborious and cost-intensive investigations to facilitate a positive reaction outcome based on Heck-type transformations, an additional synthetic route was developed, which will be discussed in section 3.3.3.3.

3.3.3.3 Syntheses towards a photoswitchable unnatural nucleoside using Suzuki cross coupling reactions as key steps

Different research groups have reported successful assembly of asymmetric diarylethene-based scaffolds by Suzuki coupling approaches.[279,295,296,405] Compared to the Friedel-Crafts-based approach presented in section 3.3.2 or the comfortably short route employing Heck-type reactions with cyclic olefins shown in section 3.3.3.2, a synthetic procedure involving Suzuki reactions is of higher complexity. Intermediately boronic acid derivatives or boronates of building blocks have to be prepared. Isolation of such as discrete compounds or *in situ* production thus generally requires additional efforts. Pd(0)-catalyzed couplings of aryl halides and organoboronic acid-derived species, however, are outstandingly versatile, reliable and thus wide-spread.[403,404,406,407] Therefore, the Suzuki reaction is a very favorable technique realizing a **TPT2**-based photoswitchable nucleoside.

An eight-step synthesis was conceived to obtain a thiophenyl- and borolane-1,2-disubstituted cyclopentene suitable for Suzuki coupling with brominated thienopyridone **82** (Figure 125).

Bromination of 2,5-dimethylthiophene (**69**) with NBS was intended to give **93**. Lithiation followed by reaction with a trialkyl borate and subsequent hydrolysis likely enables generation of thiophenyl boronic acid derivative **94**. Suzuki coupling to 1,2-dibromocyclopentene (**95**) was planned to yield **96**, which is convertible into boronic acid compound **97**. **97** in turn has high potential to be reactive towards Suzuki-type cross coupling with **TPT2** nucleobase-derived precursor compound **82**, yielding **98**.

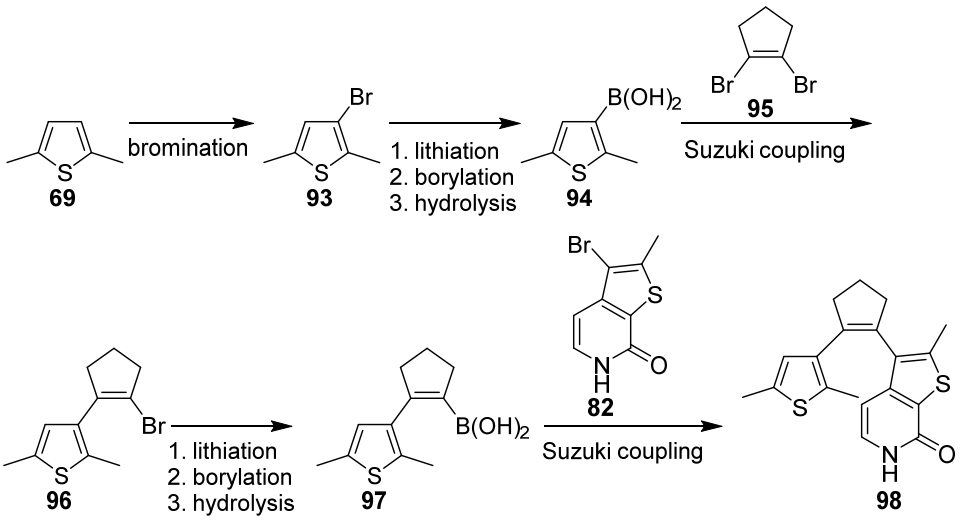


Figure 125. Proposed synthetic scheme generating boronic acid derivative **97** attached to a cyclopentenyl thiophene and subsequent attachment to **82** via Suzuki cross-coupling.

As proposed above, dimethylthiophene **69** was brominated with NBS in acetic acid with a hydroquinone additive to give **93** in a near-quantitative yield (95 %) within two hours (Figure 126).

Reaction with *n*-BuLi at -78 °C followed by the addition of tributyl borate as well as acidic workup afforded *B*-(2,5-dimethyl-3-thienyl)-boronic acid (**94**) with a 60 % yield over three steps.

Suzuki cross-coupling with 1,2-dibromocyclopentene (**95**) using a water/THF two-phase system[295] involving Palladium(0) tetrakis-triphenylphosphine as catalyst and aqueous Na_2CO_3 as base enabled isolation of 89 % **96** after 25 h of reflux.

The novel compound appeared to be rather instable and visibly decomposed within a few days at room temperature. Storage under inert atmosphere at -20 °C or lower as well as further conversion without delays was essential to facilitate the ongoing synthesis. In contrast, several similar compounds were reported by Buckup *et al*.[279] without the indication of any stability issues.

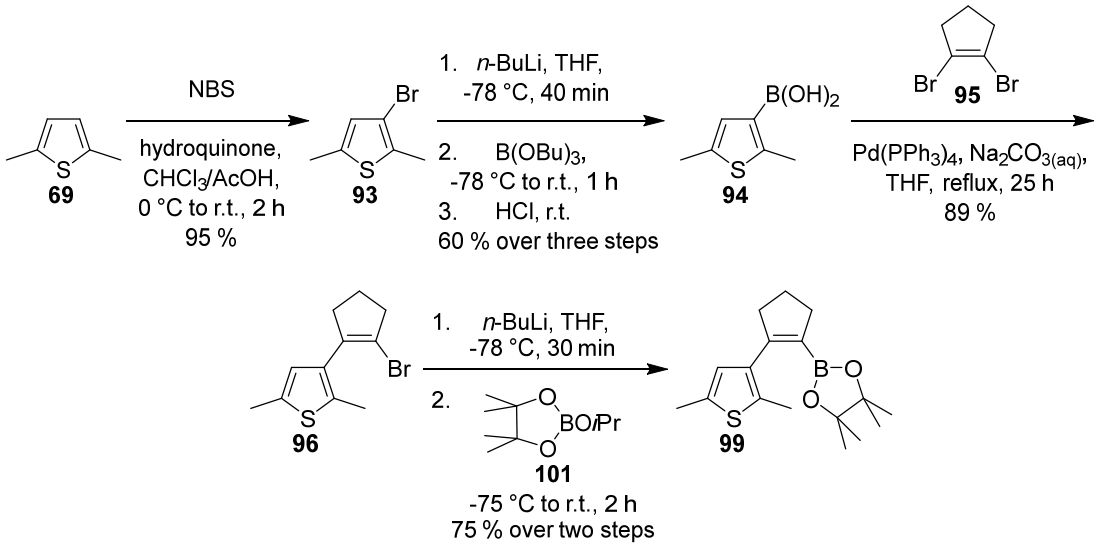


Figure 126. Synthesis of boronic acid pinacol ester derivative **99**.

Submission of **96** to a three-step approach analogous to preparation of boronic acid derivative **94** (see above) did not result in any isolatable product. It is known[85, 86] that butyl boronates of a dithienyl cyclopentene (**100**, Figure 127) deboronize during isolation; therefore, similar behavior of $B(OR)_2$-substituted variants of compound **96** is conceivable. Fortunately, **96** could be *in situ* lithiated with *n*-BuLi at -78 °C and then reacted with isopropyl boronic acid pinacol ester (**101**) yielding 75 % of **99**. This compound seemed to be stable at least for several days at room temperature under argon. Considering possible decomposition of the precious species, still straight-forward derivatization is favored in future experiments. Alternatively, boronation of **96** could be carried out *in situ* and the generated intermediate could be directly subjected to react with the brominated **TPT2** nucleobase analog precursor **82** or the peracylated ribose adduct thereof.

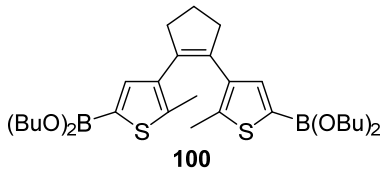


Figure 127. Chemical structure of dithiophenyl cyclopentene **100** which is labile in terms of deboronization.[85,86]

The presented procedure enables successful generation of previously unknown and versatile 2-(2,5-dimethyl-3-thienyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (**99**) with 38 % yield over seven synthetic steps. The amounts obtained were sufficient to perform further studies regarding an asymmetric dithienyl cyclopentene-based unnatural nucleoside. With primary focus on general feasibility of the conceived synthetic route only few attempts were made to optimize reaction conditions. Future works could thus improve overall yields.

Notably, the described route involved commercial 1,2-dibromocyclopentene (**95**) and in-laboratory preparation of **95** is of serious interest for future works. Synthetic procedures are reported to provide **95** in multi-gram scale from cyclopentanone (**102**) with 35–40 % overall yield.[405,410] In detail, reaction of **102** with phosphorus pentachloride intermediately gives 1-chlorocyclopentene (**103**), which can then be brominated *in situ*; eventual elimination with potassium *tert*-butoxide (*t*BuOK) facilitates conversion to **95** (Figure 128).

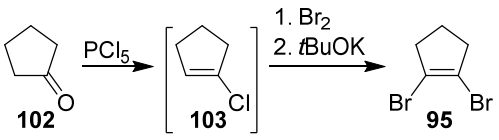


Figure 128. Synthetic scheme yielding 1,2-dibromocyclopentene (**95**) according to literature procedures.[405,410]

Commercial supply of **95** starts from ~60 € g^{-1} (*Scifinder*, accessed 2018/05); however, cyclopentanone (**102**) is available for less than 0.08 € g^{-1} (*Alfa Aesar*, accessed 2018/05). Prices of the main reagents needed for synthetic transformation to **95** are in the same range (*i.e.* PCl_5: 0.14 € g^{-1}, *t*BuOK: 0.18 € g^{-1}, and Br_2: 0.23 € g^{-1}, *abcr*/*Alfa Aesar*, requested 2019/05), making the in-house generation highly preferable.

Attachment of **82** to a protected ribose scaffold[313,316]

Previous synthetic attempts employing **82** as starting material in various reactions turned out to be challenging and partly inconclusive; the straight-forward strategy towards **62** was thus altered slightly. Fusion of the rather unreactive thienopyridone scaffold with protected ribose **37** before conduction of further steps should at least relieve from the risk of the amide function interfering with reaction success.

With the brominated nucleobase analog precursor **82** in hand, the coupling reaction to peracylated ribose **37** was performed under Vorbrüggen conditions in analogy to the procedure towards r**TPT3^{CP}** TP (**34**) described in section 3.1.1.3 (Figure 129). However, visible differences were noted during intermediate silylation with HMDS and TMS-Cl in MeCN. In previous investigations the suspension of **TPT3** nucleobase analog precursor **20** directly became clear upon addition of HMDS. In contrast, the analogous execution with **82** remained turbid even after prolonged periods of time. In various attempts proceeding with the experimental setup from the slurry did not result in product formation. The protocol was changed and the reaction mixture of **82**, HMDS and TMS-Cl was heated to reflux until it became clear, likely indicating full conversion into its trimethylsilyl ether derivative. Subsequent TFMSA-catalyzed addition to **37** was performed in a two-day reaction at room temperature.

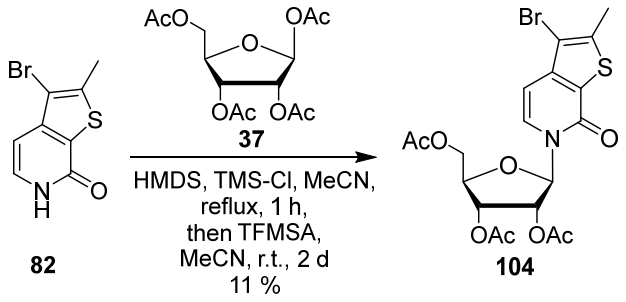


Figure 129. Synthesis of peracylated ribose adduct **104** by reaction of thienopyridone derivative **82** and **37** under Vorbrüggen conditions.

During workup a major fraction of unreacted **82** precipitated (45 % of total starting material) and could be recovered. After reversed phase HPLC purification 11 % of the 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl) adduct **104** could be obtained.

A considerable share of starting material appears to be unreacted, which might indicate a low reactivity of silylated **82** towards peracylated ribose **37** under the applied conditions. Prolonged reaction times and elevated temperatures could potentially increase yields and should be subject of further studies. Choice of solvent, catalyst, and other considerations discussed in section 3.1.1.3 could also have favorable impact on the sugar coupling efficiency. In addition, reverse phase HPLC might not be the optimal purification method. Product **104** was obtained in very high purity, however it is reasonable, that a significant fraction of the apolar compound remained undissolved (and thus is not available for purification) as the crude product is suspended in 50 % MeCN/H_2O for submission to HPLC. Application of automated flash column chromatography employing a cyclohexane/ethyl acetate (Cy/EtOAc) gradient might help increase the isolatable product fraction and is currently under investigation.

Fusion of nucleoside precursor **104** and thienyl cyclopentene derivative **99**[296]

Successful preparation of the novel nucleoside precursor **104** bearing a brominated unnatural nucleobase analog allowed for a first, low-milligram synthetic attempt to synthesize the dithienylcyclopentene compound **105**. In analogy to a promising literature procedure by Cahová and Jäschke[296] **104** and **99** were refluxed in water/acetonitrile with $Pd(OAc)_2$ serving as Pd(0) source, triphenylphosphine trisulfonate (TPPTS) as water soluble ligand and sodium carbonate as base (Figure 130).

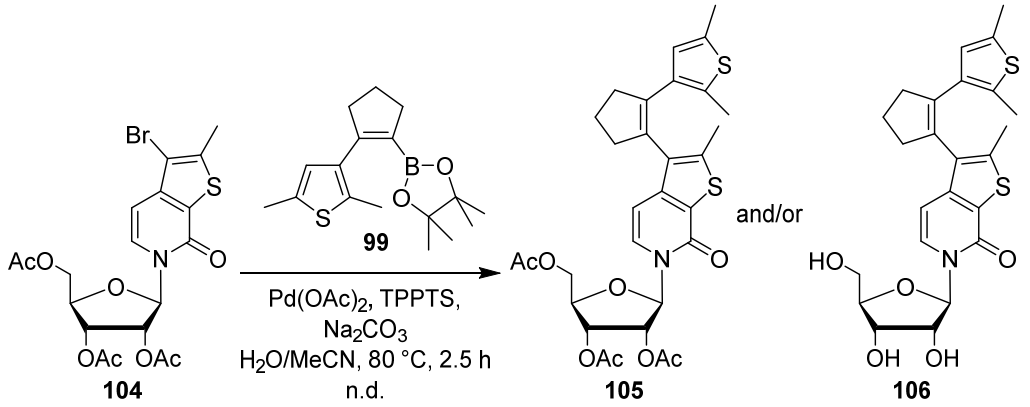


Figure 130. Suzuki coupling of unnatural nucleoside precursor **104** and **99**.

The reaction progress was tracked by HPLC-MS analysis of the complex mixture (Figure 131). A first sample was taken after 30 min reaction time and revealed formation of the desired product **105** (t_R = 14.4 min) as well as starting material **104** (t_R = 9.7 min) could also be observed. However, boronate derivative **99** could not be detected, most likely due to poor ionization properties under ESI conditions in a formic acid-supplemented mobile phase. Analysis of the coupling reaction after 2 h reflux revealed a strongly increased product peak. Surprisingly, the peak eluting at approximately 9.5 min did no longer show the mass of starting material, but the mass of acyl deprotected product species **106**. Hydrolysis of the acyl ester protective groups is conceivable under the applied reaction conditions; therefore, future studies on this project should consider altering the synthetic route to first intentionally remove acyl groups from the ribose moiety and perform Pd cross-coupling afterwards as discussed below. No starting material was detected after 2 h reaction time, thus reflux was stopped and the crude product subjected to preparative HPLC purification. Evaporation of product fractions enabled isolation of **106** in trace amounts. In contrast, **105** could not be found. Yet, it was possible to confirm the presence of deprotected nucleoside **106** by NMR and high resolution mass spectrometry. Detailed analysis of the HPLC data in retrospect further revealed existence of acyl deprotected starting material.

Altogether, these encouraging findings verify feasibility to fuse **104** and **99** under Suzuki conditions and prospective works will definitively lead to successful generation of a novel and potentially photoswitchable unnatural nucleoside.

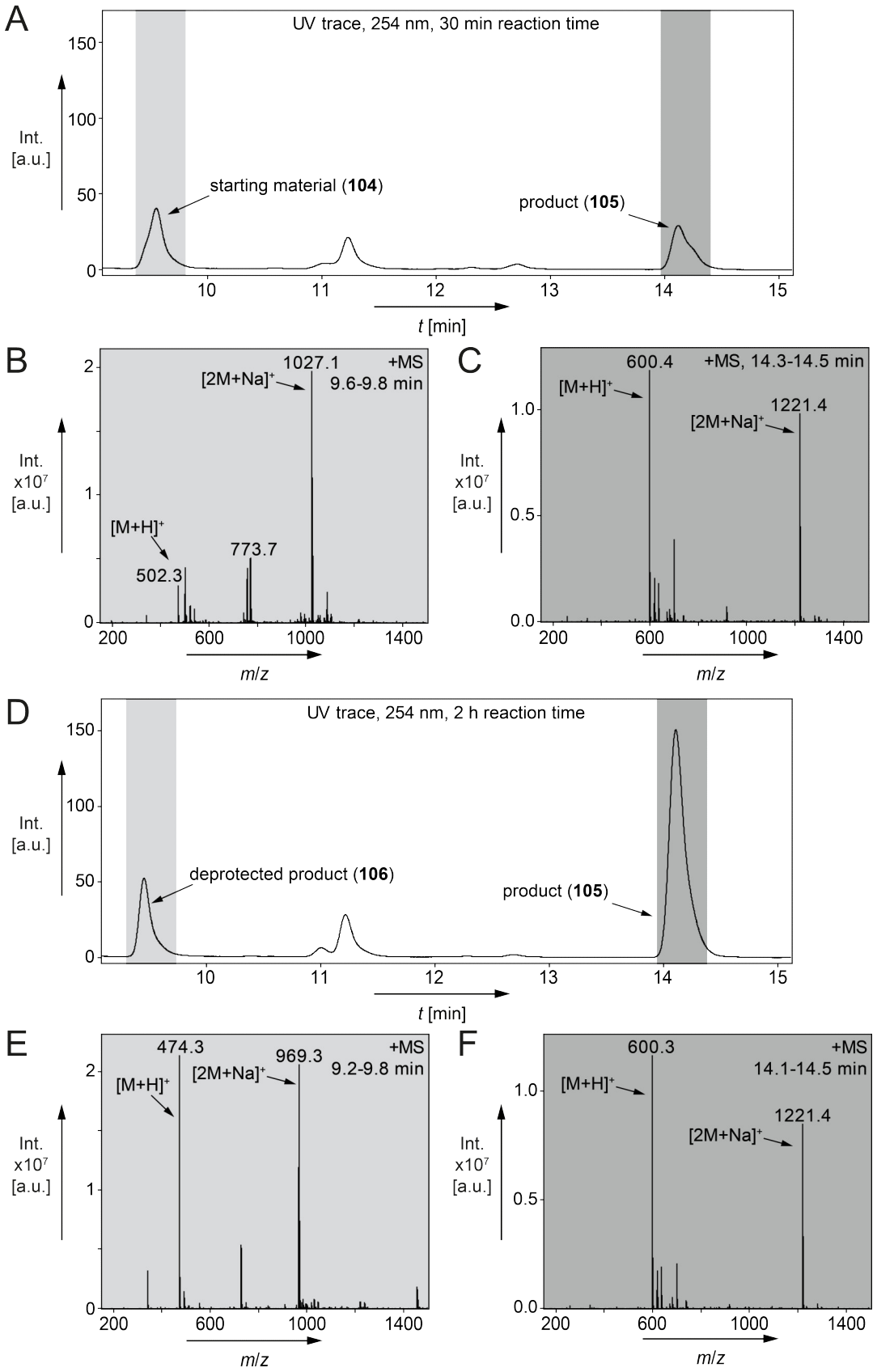


Figure 131. HPLC-MS analysis (20→100 % MeCN/0.1 % formic acid in 20 min, 0.4 mL min^{-1}, column *Zorbax*, see 5.2.4) of a Suzuki coupling reaction between **104** and **99**; A. UV trace after 30 min reaction time; B. ESI$^+$ spectrum of starting material **104** (t_R = 9.7 min, $M_{calcd.}$ = 502.0 for $[M+H]^+$); C. ESI$^+$ spectrum of product **105** (t_R = 14.4 min, $M_{calcd.}$ = 600.2 for $[M+H]^+$); D. UV trace after 2 h reaction time; E. ESI$^+$ spectrum of the peak at ~9.5 min (**106**, $M_{calcd.}$ = 474.1 for $[M+H]^+$); F. ESI$^+$ spectrum of product **105** ($M_{calcd.}$ = 600.2 for $[M+H]^+$).

Since it was found, that ester hydrolysis occurs under Suzuki conditions involving Na_2CO_3 as base, the proposed alteration of the synthetic route was considered as shown in Figure 132. Earlier thionation of the protected ribose adduct **104** gives **107** which results in bromine-substituted **TPT2** nucleoside **108** upon intentional deprotection which could then be coupled to boronate derivative **99** in a palladium-catalyzed manner eventually facilitating generation of the novel photoswitchable nucleoside **TPT2^{PS}** (**62**) bearing the unnatural **TPT2** nucleobase[194] analog.

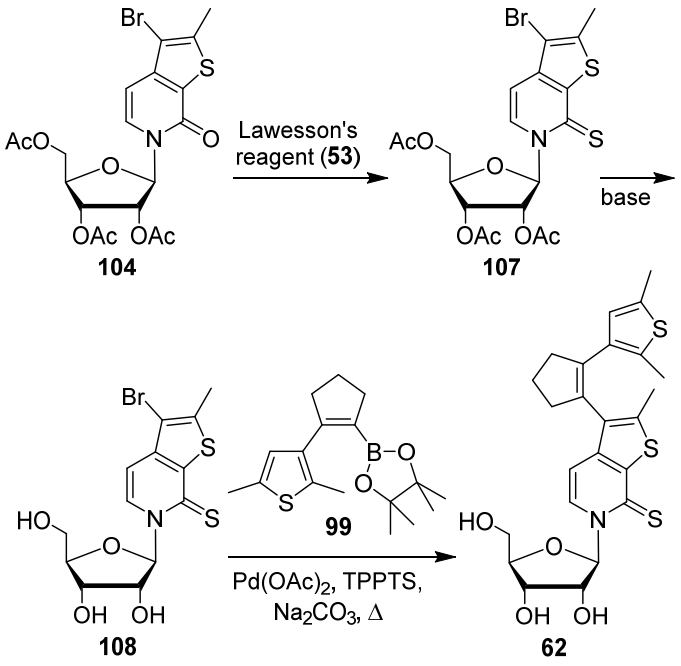


Figure 132. Proposed final synthetic steps yielding photoswitchable **TPT2**-derived nucleoside **TPT2^{PS}** (**62**).

In a first test-scale attempt thionation of **104** using Lawesson's reagent (**53**) product formation could successfully be verified (Figure 133). The reaction was followed by HPLC-MS showing serious deficits in conversion efficiency. Over an extended reaction time of 14 h refluxing in toluene, only approx. 35 % were reacted to **107** and further significant turnover was not expected. Reverse phase HPLC purification lead to recovery of unreacted starting material as well as trace amounts of the desired product. Moreover, a proton NMR spectrum and the accurate mass of **107** could be obtained to confirm the novel compound. However, no significant amounts were isolatable by the purification method. As discussed above, sample preparation for preparative HPLC might leave significant amounts of product undissolved; therefore, silica-based flash chromatography should be addressed in future experiments.

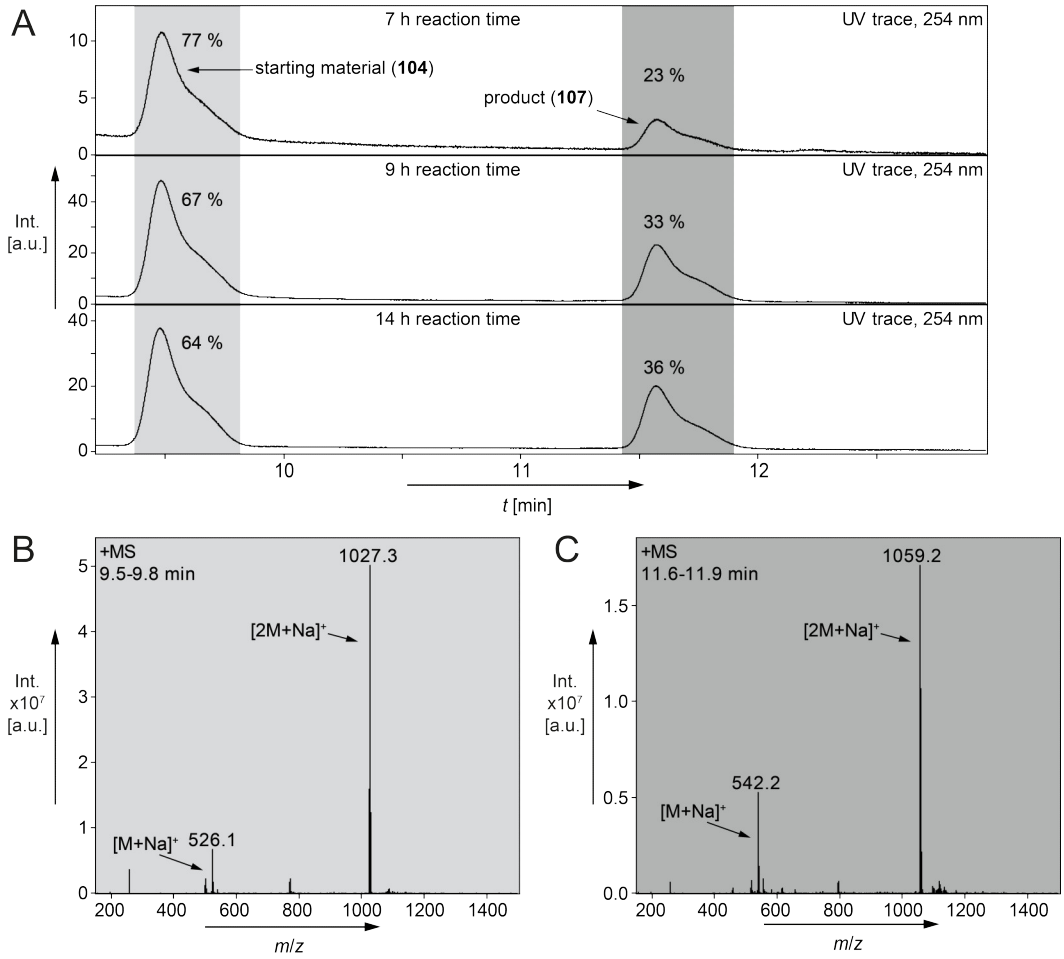


Figure 133. Thionation reaction of **104** with Lawesson's reagent (**53**) followed by HPLC-MS; A. UV traces at 7 h, 9 h, and 14 h reaction time; B. ESI$^+$ spectrum of starting material **104** (t_R = 9.6 min, $M_{calcd.}$ = 524.0 for [M+Na]$^+$) C. ESI$^+$ spectrum of product **107** (t_R = 11.7 min, $M_{calcd.}$ = 540.0 for [M+Na]$^+$).

Although further experiments towards photoswitchable unnatural nucleoside **TPT2^{PS}** (**62**) could not be performed, enormous progress was achieved establishing a straight-forward synthetic route. Several obstacles were overcome and the final approach was proven to eventually facilitate reaching the proposed goal in a short-term perspective. Completion of the presented efforts in future works can potentially enable the site-specific introduction of **TPT2^{PS}** nucleotides into RNA. Therefore, **62** could either be converted into the corresponding triphosphate to serve as cognate building block for *in vitro* transcriptions from d**NaM**-containing DNA templates, or used to build the associated cyanoethyl phosphoramidite for solid-phase RNA synthesis. Both molecules would open the fascinating possibility to control RNA-RNA, RNA-DNA, and RNA-protein interactions in a light-controlled and reversible way. Furthermore, structure formation and conformational changes of RNA could be induced, inhibited or even enhanced with time and space resolution. This technique will likely set a milestone on the way to develop new methods to analyze the biological roles of ncRNA molecules and to therapeutically address biologically relevant targets in an unprecedented manner.

3.4 Working with an expanded genetic alphabet

This study has discovered numerous new applications and insights into the art of site-specific RNA labeling using the **TPT3**:**NaM** unnatural base pair. Many works in the Kath-Schorr group during this work (*e.g.* sections 3.1.2.2, 3.1.2.3, and 3.2.2.3) and other studies crucially depend on the preparation and supply of d**TPT3** TP (**22**) and d**NaM** TP (**24**). These building blocks have been synthesized by C. Domnick[327] and as part of this thesis according to literature procedures, described in sections 3.4.1.1 and 3.4.1.2 .[188,194]

Furthermore, the preparation of useful novel unnatural base compounds is reported in sections 3.4.1.3 and 3.4.1.4. Successful synthesis of ribofuranosyl triphosphate r**TPT3^{I}** TP (**109**) enabled utilization as control substance for expanded genetic alphabet transcriptions and could serve to post-transcriptionally attach chemical functions to r**TPT3^{I}**-modified RNAs under transition metal-catalyzed cross-coupling conditions in the future.

First detailed demonstration of r**TPT3** TP (**23**) assembly allows gaining deeper insights into biochemical properties of the **TPT3**:**NaM** UBP. Nucleoside digestions coupled with HPLC analysis and development of a reverse transcription approach facilitated assessing unnatural building block incorporation efficiencies during *in vitro* transcription. These and further findings will be discussed in the following section.

3.4.1 Synthesis of unnatural nucleic acid building blocks

In order to enable PCR amplification of UBP-modified dsDNA templates (*e.g.* sections 3.1.2.2, 3.1.2.3, and 3.2.2.3) and studies on reverse transcription[370] (section 3.4.3), preparation of the deoxy triphosphates d**TPT3** TP (**22**) and d**NaM** TP (**24**) was necessary. In addition to the in-hand compound r**TPT3^{CP}** TP (**34**), the synthesis of an iodinated ribose triphosphate (r**TPT3^{I}** TP, **109**) as well as the unmodified r**TPT3** TP (**23**) were aimed. Detailed information on synthetic procedures to generate the aspired nucleotide species including highlighting of limiting steps and challenges is provided.

3.4.1.1 Synthesis of dTPT3 TP (22)

Synthesis of deoxy triphosphate d**TPT3** TP (**22**) was conducted starting from thienopyridone **20** (preparation as described in section 3.1.1.2) according to literature procedures by Li *et al*.[194] (Figure 134). Briefly, toluoyl(Tol)-protected ribose adduct **110** was obtained with 13 % yield via *in situ* silylation with HMDS and *N*,*O*-bis(trimethylsilyl)acetamide (BSA) followed by $SnCl_4$-mediated overnight reaction with chloro sugar **111**. Thionation of the pyridone carbonyl oxygen atom with Lawesson's reagent (**53**) yielded 20 % of **112**.

Subsequent removal of protective groups using sodium methoxide (NaOMe) enabled isolation of 92 % d**TPT3** nucleoside (**113**). Triphosphate synthesis resulted in 10 % d**TPT3** TP (**22**).

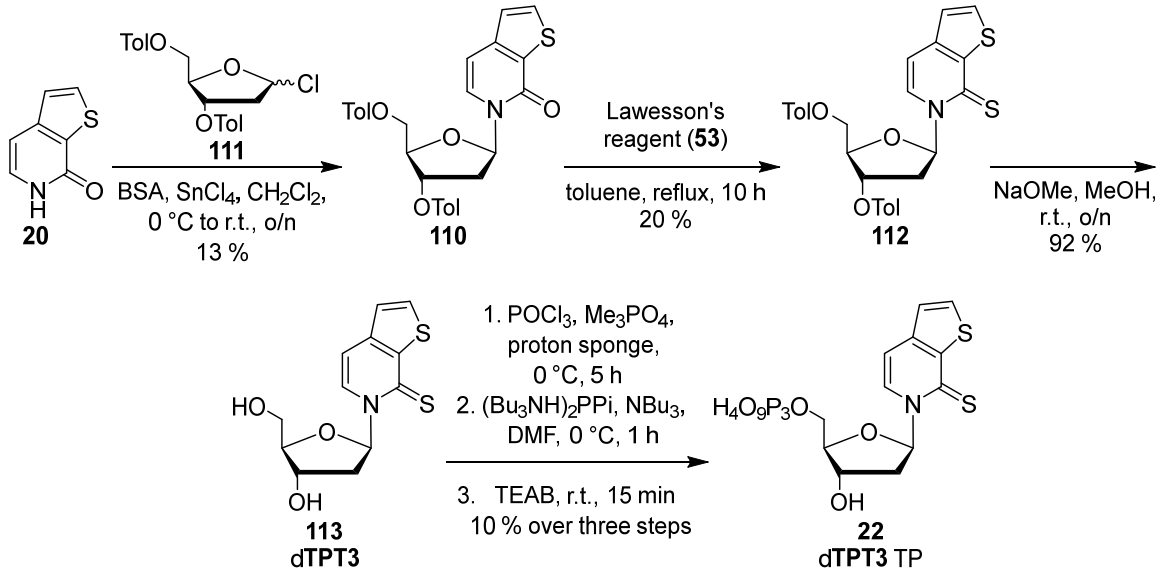


Figure 134. Synthetic procedure generating d**TPT3** TP (**22**).

Considering this six-step approach, the overall yield of 0.24 % involves many steps that are highly desirable to be improved in order to facilitate sustainable and affordable in-laboratory production of triphosphate **22**. d**TPT3** TP (**22**) is an essential compound performing enzymatic experiments involving the d**NaM**:d**TPT3** unnatural base pair on DNA level.
Notably, reproduction of the proposed and literature-based synthetic steps proved to be unexpectedly challenging, which will be discussed in detail.

Coupling of the nucleobase precursor **20** to protected deoxyribose **111**

Using reaction conditions based on the Vorbrüggen nucleosidation[316], thienopyridone **20** was intermediately silylated using *N*,*O*-bis(trimethylsilyl)acetamide in dichloromethane. Addition of toluoyl protected deoxyribosyl chloride **111** and stannic chloride enabled isolation of adduct **110** with 13 % yield after overnight agitation at room temperature. This is far below the reported literature value (39 %)[194] and can be interpreted by various facts.
Chloro sugar **111** was obtained commercially and is not classified as pure α-anomer by the supplier (*Carbosynth*, product nr. MC06108, accessed 2019/05). This might even imply a 1:1 ratio with the β-chloro species; therefore, reaction success is possibly limited to 50 %. The major route to synthesize deoxyribosyl chloride **111**, however, presents crystallization of the pure α-form.[411]

Deoxyribose **111** is also prone to degradation; even under inert atmosphere at -20 °C visible changes of the substance from clear white appearance to grey and eventually black allows estimating the decreasing quality. Literature indicates major decomposition of **111** in $CDCl_3$ within 24 h[411], stating that stability in solution is even lower.

Furthermore, it is not particularly clear that the reaction exclusively proceeds via S_N2 mechanism, thus it might actually be possible to obtain the undesired α-thienopyridinyl adduct from pure α-chloro sugar **111**.

Most prevalently hampering quantitative isolation of total product fraction is the separation from a byproduct, which is thought to be the α-nucleoside precursor. Conversion of **20** and **111** to **110** seemed to be about 50 % according to qualitative analysis of crude reactions by TLC. A difference in R_f by approx. 0.05-0.10 and the general tendency of sugar-derived compounds to elute from columns in a broad manner hindered straightforward fractioning.

Efforts establishing automated chromatography separation could help resolve this issue in future experiments.

Thionation of unnatural deoxynucleoside precursor **112**

Replacement of the pyridone oxygen atom by sulfur using Lawesson's reagent (**53**) led to 20 % of **112** after 10 h reflux in toluene followed by chromatographic purification (Lit.: 33 %[194]). Notably, the reaction tends to give quite heterogeneous mixtures of hardly separable compounds. The identity of the side products remained unresolved; however a major share is thought to be a decomposed Lawesson's reagent species. Literature does not give extensive information about side product formation, the only species discussed is trimer **114**[324] (Figure 135).

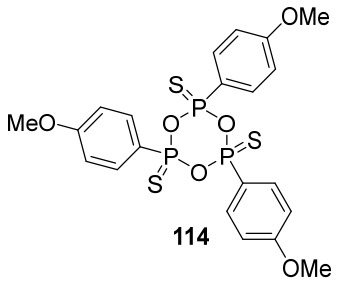


Figure 135. Literature-known[324] side product **114** of thionation reactions with Lawesson's reagent (**53**).

In many cases, a potentially Lawesson's reagent-derived byproduct was isolated. Proton NMR analysis shows advanced multiplicities in the aromatic field and two singlets around 4 ppm (Figure 136). By means of integral values, a total of twelve aromatic protons as well as nine heteroatom-bound protons could account for three phenyloxymethyl residues. In contrast the spectrum of apparent complexity is not expected for a highly symmetric compound like **114**.

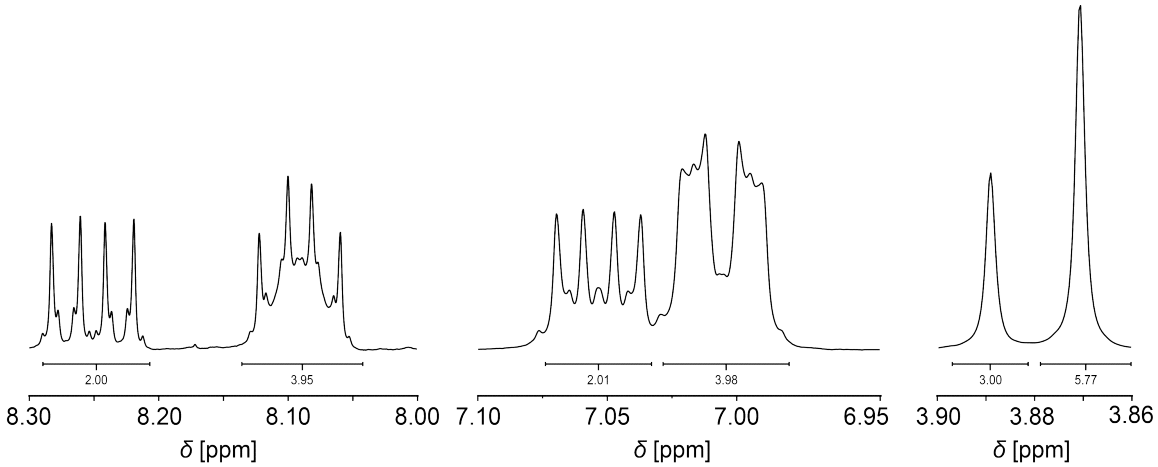


Figure 136. ^{1}H-NMR analytic sections ($CDCl_3$, 400 MHz) of an unidentified thionation side product.

Monitoring reaction progress via TLC often exhibited unexpectedly complex patterns, which may indicate formation of multiply thionated species. Potentially, ester carbonyl oxygen atoms of the toluoyl protective groups also undergo transformation with Lawesson's reagent (**53**). Notably, this has not been observed in other synthetic procedures with similar setup during this thesis. Amides are more reactive towards **53** than esters[324]; therefore, completion of thiopyridone formation is suspected upon the evidence of ester thionation. Potentially, yields could be maximized by conduction of the subsequent deprotection step without further purification. Hypothesizing quantitative pyridone thionation, removal of toluoyl groups (thionated or not thionated) would leave a single species to be isolated with ease.

Several attempts to realize successful isolation of **112** even under optimized conditions did not result in a positive outcome and a precise cause for reaction failure could not be specified. The reaction remains challenging in terms of reproducibility. However, it was possible to gain satisfactory amounts of thionated toluoyl deoxynucleoside **112** to accomplish further synthetic steps.

Removal of hydroxyl protective groups

Li *et al.* reported deprotection of toluoyl deoxyribose **112** by a one hour reaction with sodium methoxide in 82 % yield.[194] Prolonging the reaction time to an overnight reaction during the course of this work, isolation of product **113** could be achieved with 92 %. This reaction usually does not result in any inconvenience and further improvement of the procedure is not necessary.

Triphosphate synthesis

Triphosphorylation was achieved using the general procedure according to Ludwig[315]. A five hour reaction with $POCl_3$ under ice cooling followed by pyrophosphate addition and buffered hydrolysis of the intermediately formed cyclic triphosphate yielded a considerably moderate yield of 10 % d**TPT3** TP (**22**) after HPLC purification. The reported literature yield of 30 % approaches the upper limit of commonly observed reaction outcomes for this delicate procedure.[194,305] Li *et al.* used a lower reaction temperature of -10 °C for the first reaction step as well as a shorter reaction time (2 h).[194] The experience gathered throughout this work suggested significantly differing reaction times for the 5′-monophosphorylation depending on the nucleoside; therefore, prolonged reaction times are recommended to ensure maximum conversion. Fine tuning of cooling conditions might have considerable impact as the risk of byproducts (additional 3′-phosphorylation for deoxy nucleosides and 2′,3′-cyclic phosphate formation for ribose nucleosides, respectively)[412] might increase with higher temperatures.

d**TPT3** TP (**22**) could be isolated in high purity as confirmed by HPLC-MS analysis (Figure 137). For enzymatic applications such as PCR amplification involving the d**TPT3**:d**NaM** base pair the highest quality of triphosphates is required, thus careful purification is essential.

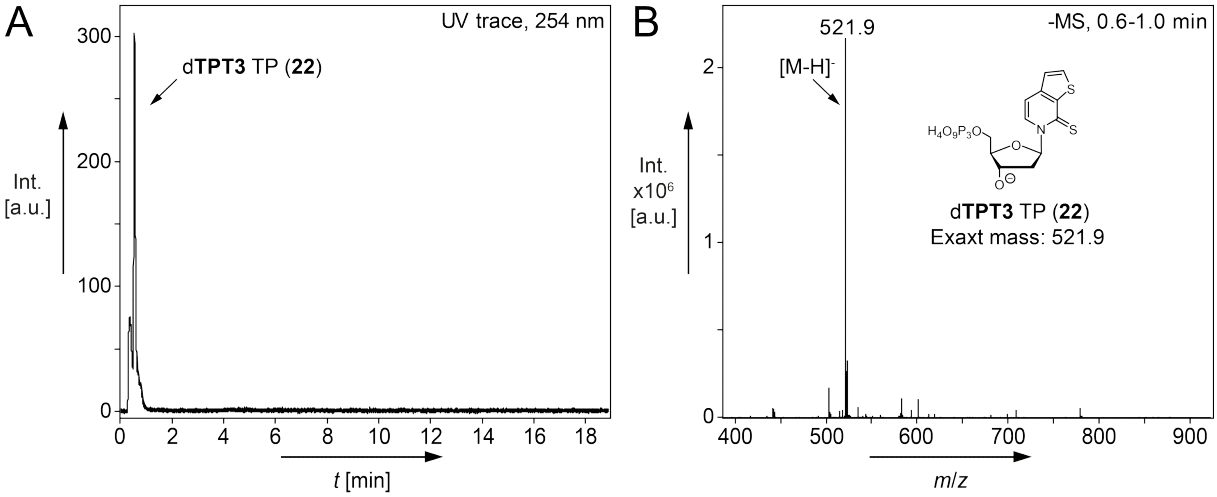


Figure 137. HPLC-MS analysis of purified d**TPT3** TP (**22**) (see 5.2.4 for details); A. UV trace; B. ESI^- spectrum of **22** (t_R = 0.8 min, $M_{calcd.}$ = 521.9 for $[M-H]^-$).

With **22** one part of the deoxy nucleotide building blocks for expanded genetic alphabet experiments on DNA level was successfully prepared, production of the counterpart triphosphate d**NaM** TP (**24**) is shown in the following section.

3.4.1.2 Synthesis of dNaM TP (24)

Nucleoside d**NaM** (**115**) is commercially available from *Berry & Associates, inc*. (174 $ per 25 mg, accessed 2019/05) in high purity. The convenience and reliability of the product purchase relieves from a laborious synthesis towards **115**. During the studies of C. Domnick[327] in the Kath-Schorr group preparation of d**NaM** (**115**) could be achieved with approx. 5 % overall yield in a four-step approach starting from deoxyribose and 2-methoxynaphthalene.

Triphosphate synthesis was carried out according to standard procedures as provided in the literature (Figure 138).[315] As mentioned above, tendencies of phosphorylation kinetics using $POCl_3$ were experienced to be considerably individual. 6 h reaction time under ice cooling should be sufficient in most cases. Addition of pyrophosphate might be the critical step (see section 3.1.1.3); therefore, reaction time with bis(tributylammonium) pyrophosphate was prolonged from literature values of 5 min to a 1 h reaction. d**NaM** TP (**24**) could be obtained with a yield of 4 %, still representing a reasonable range for delicate triphosphate synthesis procedures in such small scale. The initial procedure described by Seo *et al.* does not state any yield.[188] C. Domnick could accomplish a maximum isolation of 11 % **24**.[327]

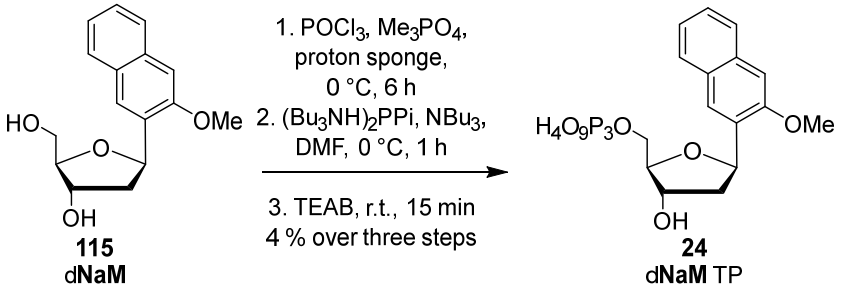


Figure 138. Synthesis of d**NaM** TP (**24**).

Successful quality analysis was performed employing HPLC-MS (Figure 139).

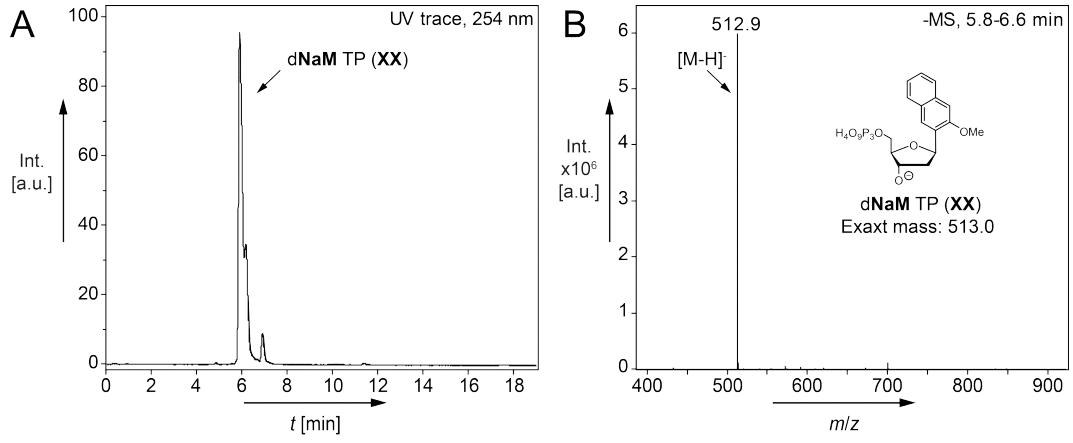


Figure 139. HPLC-MS analysis of purified d**NaM** TP (**24**) (see 5.2.4 for details); A. UV trace; B. ESI^- spectrum of **24** (t_R = 6.0 min, $M_{calcd.}$ = 513.0 for $[M-H]^-$).

With generation of both deoxyribonucleoside triphosphates of the **TPT3**:**NaM** base pair it was possible to conduct various DNA-based applications involving this UBP. Lacking commercial availability, constant supply with **22** and **24** was necessary (as also performed by C. Domnick[327]). By these efforts, ongoing research progress was ensured as the triphosphate building blocks of d**TPT3** and d**NaM** were *e.g.* needed in this study to generate mid-range and long dsDNA templates for *in vitro* transcription via PCR with an expanded genetic alphabet (see sections 3.1.2.2 and 3.1.2.3).

3.4.1.3 Synthesis of iodinated rTPT3^{I} TP (109)

In addition to r**TPT3**^{CP} TP (**34**), an intermediate compound of the synthetic route (section 3.1.1.3), iodinated unnatural ribose nucleoside **51** was also converted into its corresponding triphosphate compound.

Employing analogous procedures as described before, r**TPT3**^{I} nucleoside (**51**) was submitted to react with phosphorous oxychloride at low temperature for several hours and then subjected to bis(tributylammonium) pyrophosphate ($(Bu_3NH)_2PPi$) addition under basic conditions for 15 min (Figure 140). Hydrolysis of the formed cyclic triphosphate and reverse phase HPLC purification allowed reception of the desired triphosphate compound r**TPT3**^{I} TP (**109**) in 8 % yield.

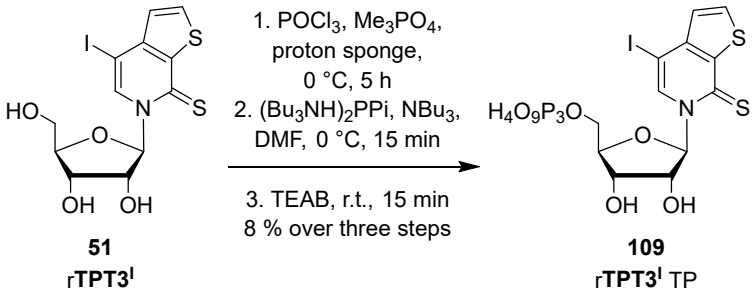


Figure 140. Triphosphate synthesis yielding r**TPT3**^{I} TP (**109**).

In addition to standard verification methods ($^{1}H/^{31}P$-NMR, high resolution (HR) MS) purity and integrity was analyzed via HPLC-MS (Figure 141). Both the UV trace and the raw ESI^- data confirm high quality of the isolated triphosphate r**TPT3**^{I} TP (**109**).

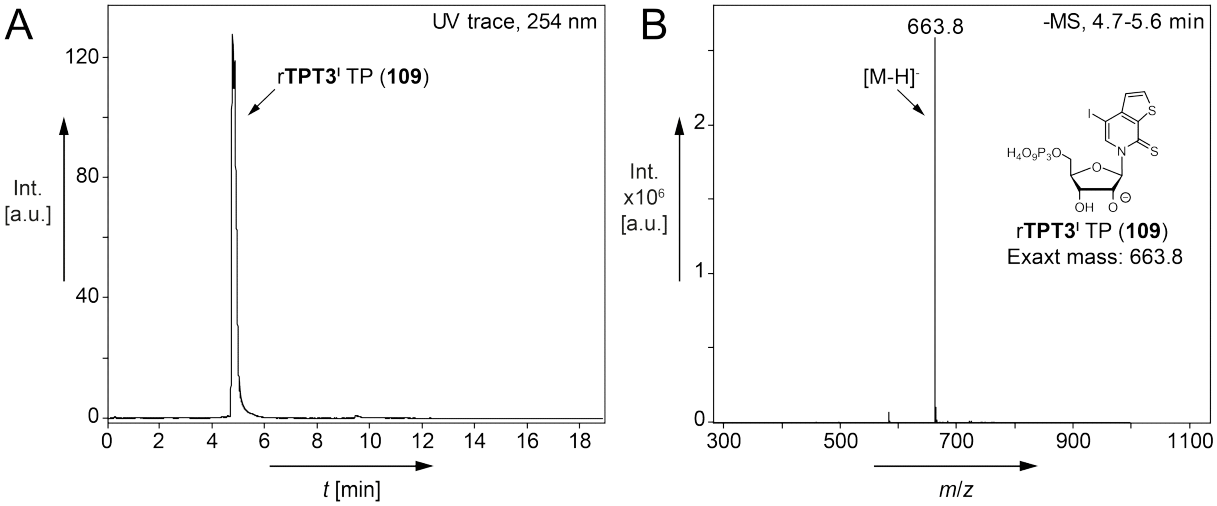


Figure 141. HPLC-MS analysis of purified r**TPT3^{I}** TP (**109**) (see 5.2.4 for details); A. UV trace; B. ESI^{-} spectrum of **109** (t_R = 5.2 min, $M_{calcd.}$ = 663.8 for [M-H]$^{-}$).

The unnatural ribonucleotide **109** has high potential to serve as valuable tool in various contexts. For example, it can serve as model substance to study feasibility and efficiency of template-directed incorporation into RNA via T7 *in vitro* transcription from d**NaM**-containing DNA. Due to the comparably large iodine atom substituent **109** could be used to estimate incorporation of r**TPT3**-derived molecules bearing linker systems at the same position.

When short unnatural base-modified RNA molecules suitable for mass spectrometric analysis are generated, application of r**TPT3^{I}** nucleotides has the advantage to exhibit a distinct atomic mass which is easily distinguishable from *e.g.* unmodified sequences. *In vitro* transcription from d**NaM**-containing DNA templates results in efficient incorporation into RNA as can be seen from HPLC-MS analysis of a crude transcription reaction from **DNA^{NaM}** in the presence of r**TPT3^{I}** TP (**109**) yielding **RNA^{I}** (Figure 142). Triphosphate **109** is thus frequently used as an internal standard in the Kath-Schorr group giving excellent feedback on expanded genetic alphabet transcription reactions.

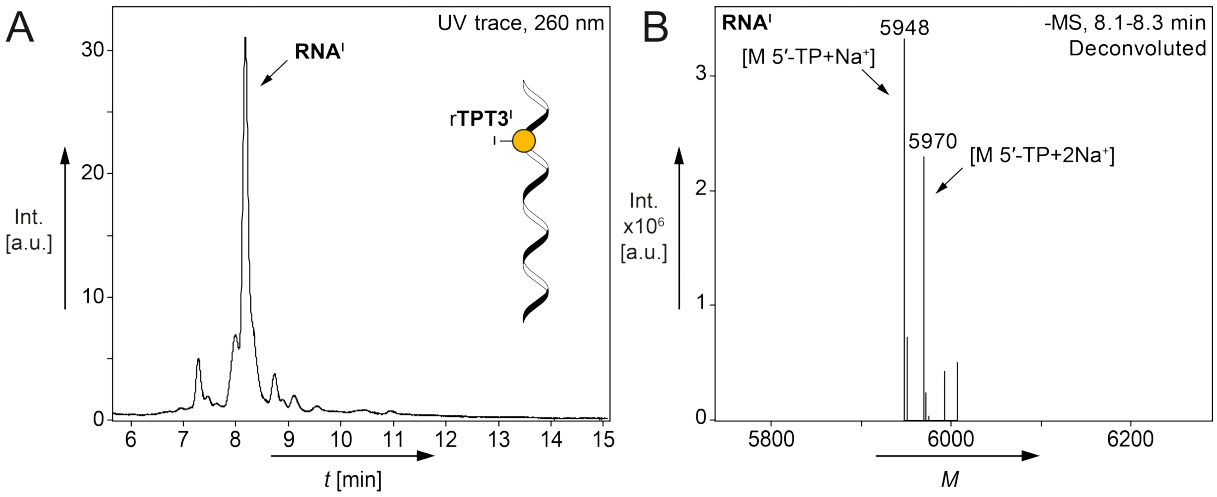


Figure 142. HPLC-MS analysis of a crude T7 *in vitro* transcription from **DNA^{NaM}** with r**TPT3^{I}** TP (**109**) (see 5.3.13 for details); A; UV trace; B. Deconvoluted ESI^{-} mass spectrum of the peak eluting at t_R ~8.2 min ($M_{calcd.}$ = 5948 for [M 5′-TP+Na^{+}], $M_{calcd.}$ = 5970 for [M 5′-TP+2Na^{+}]).

Furthermore, triphosphate **109** could prospectively be utilized to post-transcriptionally functionalize transcribed RNA containing r**TPT3^{I}** nucleotides employing boronate derivatives in aqueous media-based Suzuki reactions. Literature has proven biocompatible Suzuki-Miyaura cross-coupling to be a highly important transformation attaching functional probes under mild conditions. For example, Cahová and Jäschke were able to generate photoswitchable entities on synthetically synthesized DNA oligonucleotides containing 5-iodo-2′-deoxyuridine using Suzuki reactions to reversibly control transcription efficiency thereof.[296] Furthermore, Lercher *et al*. improved this reaction strategy and effectively reacted 5-iodo-2′-deoxyuridine-modified DNA with a variety of functional probes demonstrating the versatility of this approach.[413] Walunj *et al*. reported successful attachment of fluorogenic reporters to *in vitro* transcribed RNA with 5-iodouridine nucleotides under Pd-catalyzed cross-coupling conditions, verifying the applicability of such reactions to ribonucleic acid oligomers.[129]

Site-specific introduction of r**TPT3^{I}** nucleotides into RNA would allow expanding the scope of modifications by a broad spectrum of reporter groups and will thus be of great impact in prospective research projects within the Kath-Schorr group and elsewhere.

3.4.1.4 Synthesis of rTPT3 TP (23)

Preparation of the unfunctionalized unnatural ribose nucleoside triphosphate r**TPT3** TP (**23**) was envisaged. Successful synthesis of **23** would enable enzymatic incorporation of r**TPT3** nucleotides into RNA. Comparative studies with other functionalized r**TPT3** nucleotides (*e.g.* r**TPT3^{CP}** TP (**34**), r**TPT3^{NO}** TP (**61**), or r**TPT3^{I}** TP (**109**)) regarding their respective performance in RNA modification would be facilitated. In this context, **23** would particularly

be useful as an internal standard for site-specific RNA labeling using template-directed T7 *in vitro* transcription of d**NaM**-modified DNA. In connection with r**NaM** TP (**25**), which was prepared by C. Domnick[327] in the research group of Dr. S. Kath-Schorr, triphosphate **23** could serve to gain previously inaccessible information on properties, effects, and dynamics of the r**TPT3**:r**NaM** base pair in ribonucleic acid molecules and hybrids. Thus, r**TPT3** TP (**23**) is of high and general scientific importance.

Despite being already used by the Romesberg group[198,201,203], the synthetic route had not been reported as nucleoside and triphosphate syntheses were outsourced to commercial suppliers.

Preparation of r**TPT3** TP (**23**) was envisioned to follow a similar strategy as employed for r**TPT3^{CP}** TP (**34**) (see 3.1.1.3). In this case, direct thionation of unnatural nucleoside precursor **48** was intended to give **116** Deprotection yielding nucleoside **117** and triphosphorylation were planned to easily obtain r**TPT3** TP (**23**) (Figure 143).

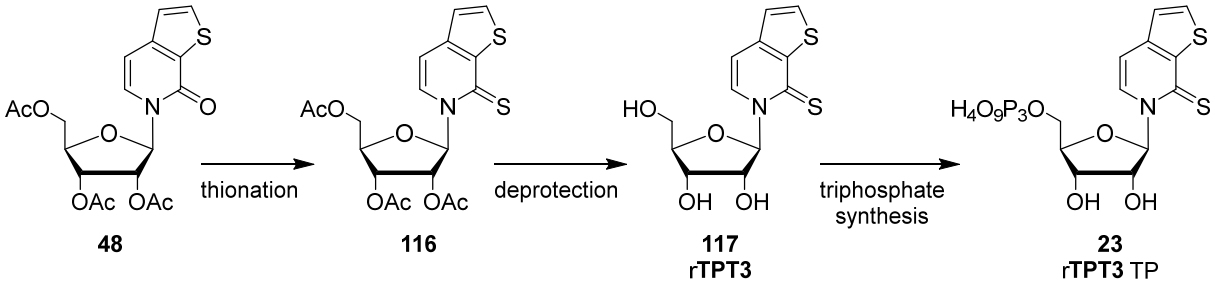


Figure 143. Schematic synthesis of r**TPT3^{TP}** (**23**).

The acyl protected ribose adduct **48** was already in hand as an intermediate compound on the synthetic route to r**TPT3^{CP}** TP (**34**) (section 3.1.1.3). Completion of the nucleobase analog via thionation with Lawesson's reagent (**53**) led to 12 % of compound **116**. Ribose acyl groups could quantitatively be removed to unnatural nucleoside **117** by a two-day reaction with methanolic ammonia (Figure 144). Triphosphate synthesis applying Ludwig conditions[315] gave, after preparative reverse phase HPLC purification, r**TPT3** TP (**23**) in a satisfying yield of 24 %. Starting from **48** the achieved overall yield is 2.9 %.

Drawbacks from the thionation reaction apply as discussed earlier in this thesis (see sections 3.1.1.3, 3.3.3.3 and 3.4.1.1).

Removal of ribose acyl groups again proceeded smoothly, no further optimization is required. However, an overnight reaction might be insufficient for complete conversion; therefore, a prolonged incubation time of two or more days is recommended to maximize product isolation.

The triphosphate synthesis employing **117** as starting material gave a surprisingly good yield compared to other procedures described throughout this thesis. No major differences were made conducting the triphosphorylation, further suggesting that these reactions are very

sensitive and remain rather hardly predictable when lower milligram scales are handled. Monophosphorylation is considered to be the major byproduct as similarly discussed in section 3.1.1.3. Also, purification and multiple transfer steps might be significant sources of yield decrease.

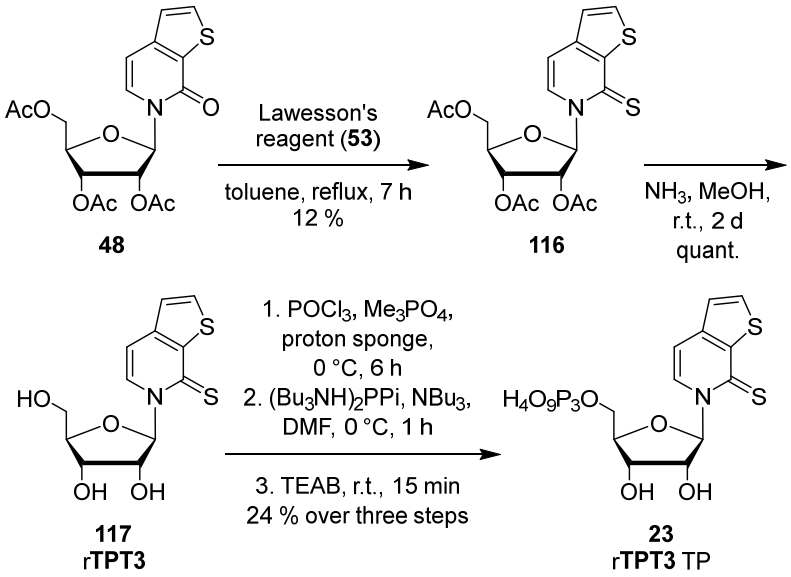


Figure 144. Synthesis of r**TPT3** TP (**23**).

Besides proton NMR, phosphorous NMR, and high resolution mass spectrometric analysis, the purity and integrity of purified r**TPT3** TP (**23**) could be confirmed by HPLC-MS analysis as shown in Figure 145.

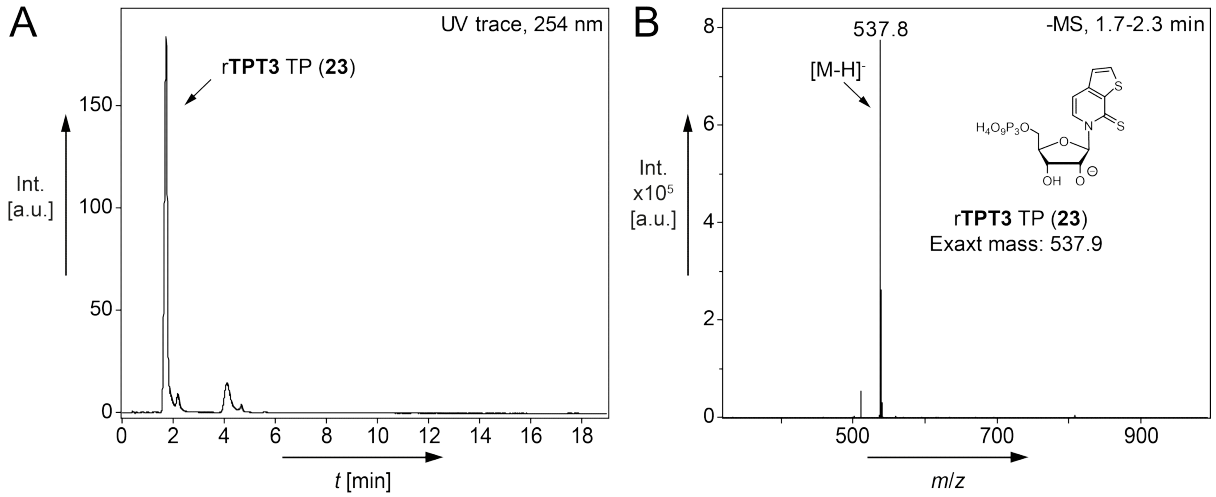


Figure 145. HPLC-MS analysis of purified r**TPT3** TP (**23**) (see 5.2.4 for details); A. UV trace; B. ESI$^-$ spectrum of **23** (t_R = 1.8 min, $M_{calcd.}$ = 537.9 for [M-H]$^-$); the minor peak eluting at ~4 min corresponds to the r**TPT3** 5′ MP and did not negatively affect transcription efficiency.

As proof of concept, triphosphate **23** was submitted to a standard *in vitro* transcription reaction from the short d**NaM**-containing test strand **DNA^{NaM}**. HPLC-MS analysis of a crude, DNase-digested and gel-filtrated sample revealed an unexpectedly homogeneous formation of the desired oligoribonucleotide **RNA^{TPT3}** (Figure 146).

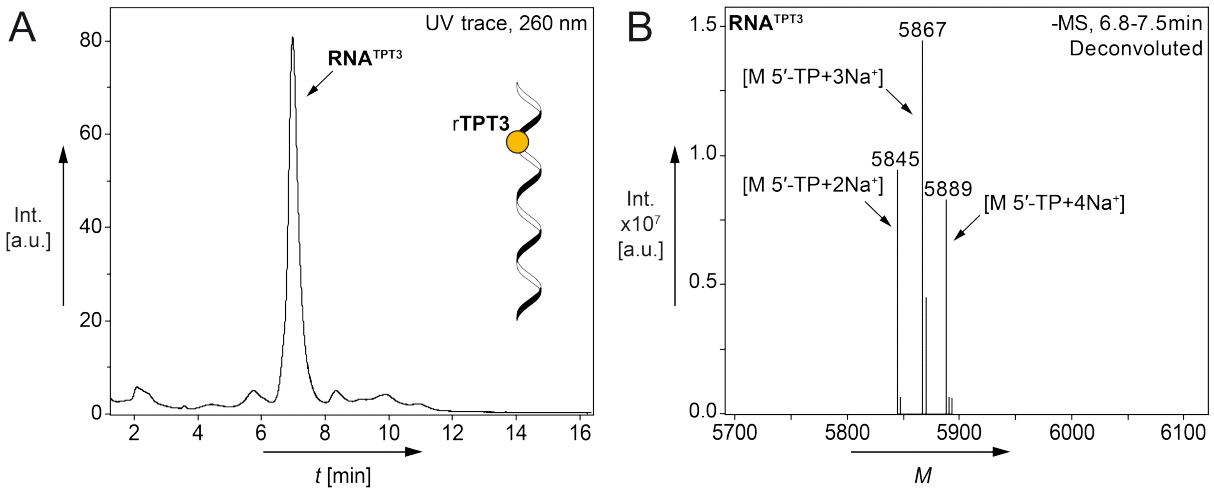


Figure 146. HPLC-MS analysis of a crude T7 *in vitro* transcription from **DNA^{NaM}** with r**TPT3** TP (**23**) (see 5.3.13 for details); A. UV trace; B. Deconvoluted mass spectrum of the peak eluting at t_R ~7 min ($M_{calcd.}$ = 5844 for [M 5′-TP+2Na$^+$], $M_{calcd.}$ = 5866 for [M 5′-TP+3Na$^+$], $M_{calcd.}$ = 5888 for [M 5′-TP+4Na$^+$]).

These results indicate excellent incorporation efficiency of T7 RNA polymerase during template-directed transcription from DNA including d**NaM** nucleotides in the presence of r**TPT3** TP (**23**).

Introduction of r**TPT3** nucleotides into RNA holds several important opportunities working with an expanded genetic alphabet based on the **TPT3**:**NaM** unnatural base pair. Options range from quantitative accessibility of transcription yields in comparison to linker-modified variants such as r**TPT3^{CP}** TP (**34**) or r**TPT3^{NO}** TP (**61**) to various fundamental research aspects investigating the **TPT3**:**NaM** unnatural base pair on RNA level.

Interesting insights into transcription efficiency and applicability of the UBP **TPT3**:**NaM** have recently been gained in the Kath-Schorr group based on a transcription-reverse transcription approach[370] and will be discussed in section 3.4.3.

3.4.2 Nucleoside digestions with an expanded genetic alphabet

Introduction of complete r**TPT3**$^{(R)}$:r**NaM** base pairs into RNA is of major interest for future implications with an augmented genetic code. Faithful and efficient incorporation of r**TPT3** derivatives into RNA by template-directed T7 *in vitro* transcription has been thoroughly demonstrated during this study (sections 3.1.2 and 3.2.2). However, detailed studies of r**NaM** nucleotides for RNA modification remain elusive.

During the course of C. Domnick's studies, HPLC-MS analyses of transcripts putatively containing r**NaM** proved to be heterogeneous. Unfortunately, the **NaM** nucleobase exhibits a closely similar mass compared to sodium adducts of **A** and **U**. Thus, unambiguous assignment of desired transcripts was not possible within the limits of accessible MS accuracy. However, it was hypothesized that r**NaM** incorporation might be as low as 16 %.[327]

Nucleotide composition of transcribed RNAs can be assessed by an enzymatic degradation approach coupled to HPLC analysis.[414] First attempts to decipher r**NaM**-containing transcripts with this method by C. Domnick were adapted and fine-tuned in this study to get insights into expanded genetic alphabet transcripts.

Notably, d**TPT3** CEP is obligatory to synthesize template DNAs for *in vitro* incorporation of r**NaM** nucleotides into RNA and has not been reported yet. However, d**5SICS** CEP is available through *Berry & Associates Inc.* (Dexter, MI, USA) and the nucleobase analog **5SICS** has been found to direct introduction of r**NaM** via transcription by the Romesberg group.[187] Thus, a template sequence, **DNA**$^{\mathbf{5SICS}}$, was designed in analogy to **DNA**$^{\mathbf{NaM}}$ (section 3.1.2.1) and provided by custom DNA synthesis (*Ella Biotech*, Martinsried, Germany). Transcription from **DNA**$^{\mathbf{5SICS}}$ in the presence of r**NaM** TP (**25**), prepared by C. Domnick[327], was intended to yield **RNA**$^{\mathbf{NaM}}$.

For analysis, procession of RNA transcripts by S1 nuclease, alkaline phosphatase, and snake venom phosphodiesterase was planned, facilitating detection of single nucleosides via their unique HPLC retention times (Figure 147). In this work, comparison of **RNA**$^{\mathbf{TPT3}}$, **RNA**$^{\mathbf{NaM}}$, and unmodified **RNA**$^{\mathbf{n1}}$ digests was intended. These sequences have identical nucleotide compositions differing only in one (unnatural) nucleotide. For further proof of reproducibility, triphosphate mixes in identical ratios to the transcript compositions were prepared and enzymatically dephosphorylated before HPLC analysis.

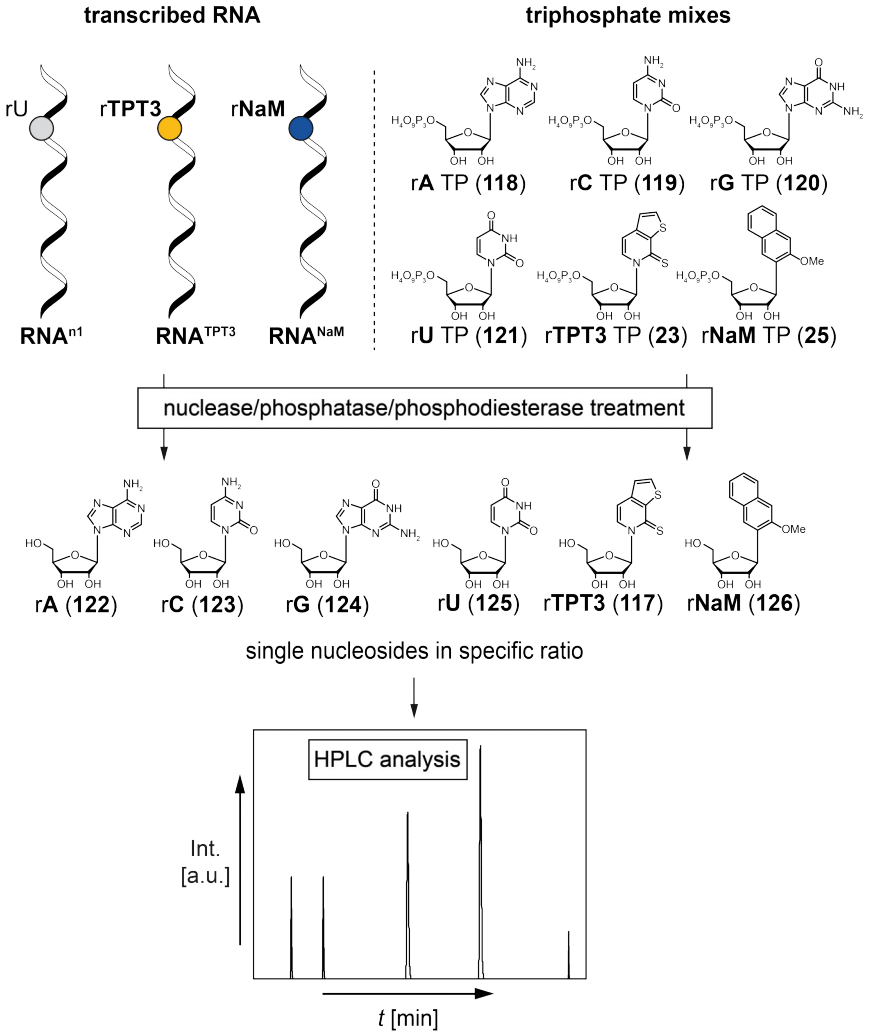


Figure 147. Schematic nucleoside digestion approach of transcribed RNAs (native **RNA^{n1}**, or unnatural base-modified **RNA^{TPT3}**/**RNA^{NaM}**) and control triphosphate mixes and detection by HPLC.

Applying an improved gradient on an alternative HPLC column (see 5.3.8), first the retention times of single nucleosides were analyzed to ensure proper separation and identification (Figure 148). 800 pmol of ribosyl triphosphates (r**A** TP (**118**), r**C** TP (**119**), r**G** TP (**120**), r**U** TP (**121**), r**TPT3** TP (**23**), and r**NaM** TP (**25**)) were treated with alkaline phosphatase and subjected to HPLC analysis after enzyme denaturation and centrifugation. The nucleosides are perfectly resolved and show unambiguous as well as reproducible retention times (t_R(r**A**, **122**) = 12.19±0.06 min, t_R(r**C**, **123**) = 2.37±0.02 min, t_R(r**G**, **124**) = 9.37±0.10 min, t_R(r**U**, **125**) = 3.48±0.04 min, t_R(r**TPT3**, **117**) = 20.72±0.04 min, t_R(r**NaM**, **126**) = 25.79±0.02 min, each value is the average of at least three independent measurements, errors are SDs thereof, see 5.3.8 for single values). The unnatural nucleosides r**TPT3** (**117**) and r**NaM** (**126**) elute at much higher retention times compared to canonical nucleosides due to their hydrophobic nature. Notably, absorption intensity decreases from purine nucleobases to pyrimidine nucleobases to unnatural nucleobase analogs. The least intensity is observed for r**NaM** (**126**).

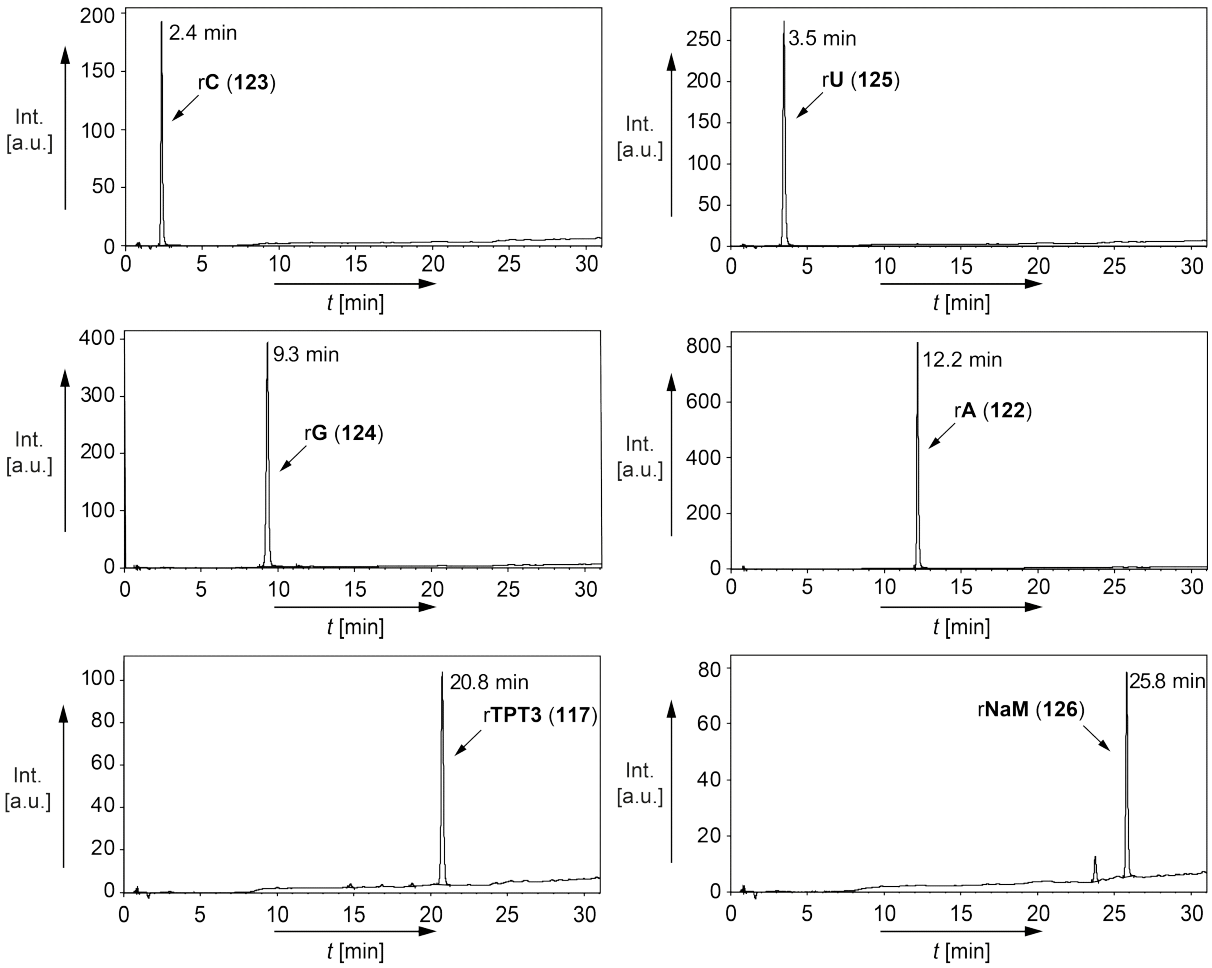


Figure 148. UV traces of phosphodiesterase-treated single ribonucleoside triphosphates analyzed via HPLC.

Next, *in vitro* transcribed RNAs (**RNA^{n1}**, **RNA^{TPT3}**, and **RNA^{NaM}**) were prepared, gel-filtrated, and submitted to multi enzyme-mediated degradation. In parallel, triphosphate mixes representing the number of nucleotides contained in the sequences were enzymatically dephosphorylated yielding nucleosides. In a comparative way, these nucleoside solutions obtained either from transcribed RNA or triphosphate mixes were subjected to HPLC separation (Figure 149). The control spectrum of **RNA^{n1}** is perfectly resembled by the corresponding digestion from the triphosphate mix. The UV trace obtained from degraded **RNA^{TPT3}** is in complete agreement to its artificially composed spectrum, clearly stating the r**TPT3** nucleotide being incorporated in the parental nucleic acid and allowing a rough estimation of a near-quantitative abundance compared to intensity of r**TPT3** (**117**) in the triphosphate mix. In contrast, **RNA^{NaM}** digested to nucleosides does not reveal doubtlessly interpretable HPLC data. The triphosphate gives a minor r**NaM** (**126**) peak which is not distinguishable from background noise in the degraded RNA spectrum. Unfortunately, even with the control data available, it cannot be assured if significant quantities of r**NaM** nucleotides were successfully incorporated in the corresponding RNA.

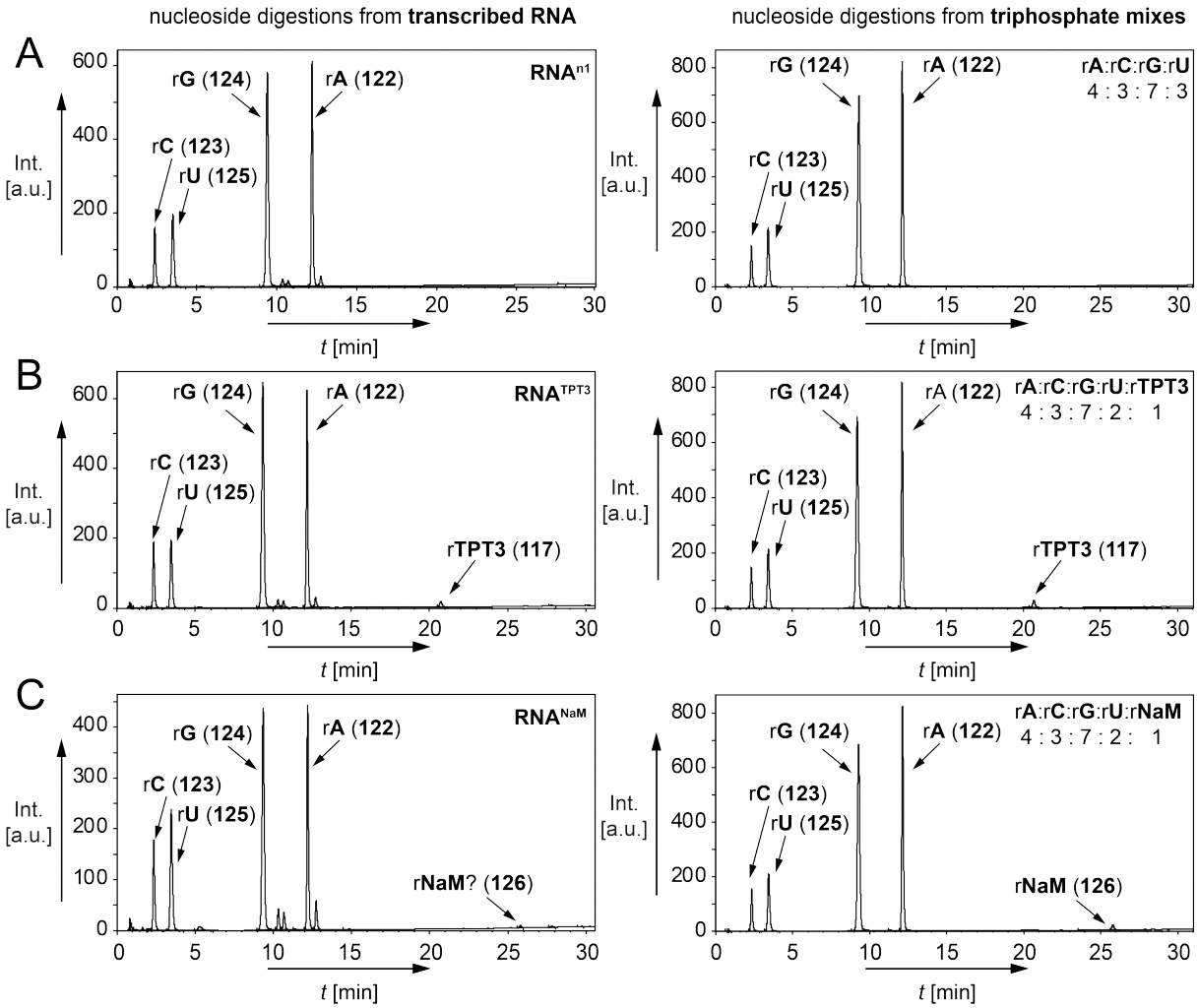


Figure 149. HPLC analyses of nucleoside digestions from transcribed RNAs (left panels) and control mixes prepared from triphosphates (right panels); A. **RNA^{n1}**; B. **RNA^{TPT3}** C. **RNA^{NaM}**.

The presented technique is a useful tool to analyze the composition of small RNAs, potentially even allowing for quantification by integrals of the UV peaks if sufficient material is applied. As a future advance, r**TPT3**-derived nucleosides could also be detected specifically due to an absorption maximum at 360 nm[392]. On the other hand, the poor absorption properties of the **NaM** nucleobase analog hinder efficient analysis. For clarity of data evaluation, vast amounts of valuable RNA would have to be sacrificed while increasing the risk of rising background signals at the same time. In addition, control experiments with synthetic RNA would be extremely useful to verify the recorded spectra, however for this the potentially challenging syntheses of r**NaM** and r**TPT3** CEP would have to be developed. The nucleoside digestion-HPLC approach might also be limited to rather short constructs in order to obtain reasonable signal ratios with the canonical nucleosides. To overcome these disadvantageous features, a new evaluation method involving fluorogenic reverse transcription was developed and is discussed in the following.

3.4.3 Reverse transcription with an expanded genetic alphabet

During the course of this study, a reverse transcription (rtc) approach applying unnatural base-modified RNA was developed and established in close collaboration with K. Kurscheidt and E. Hoffmann,.[370–372] With this technique, investigation of *in vitro* transcribed sequences with respect to site-specificity and abundance of introduced nucleotides belonging to the **TPT3**:**NaM** UBP was intended. Moreover, successful generation of a modified reverse transcript (*i.e.* complementary DNA, cDNA) with the unnatural base information retained would set a milestone towards RNA-based SELEX involving unnatural bases.

On DNA level, efforts involving an expanded genetic alphabet for the selection of aptamers (ExSELEX) resulted in remarkable improvements, *e.g.* significantly enhanced affinities and increased stabilities of aptamer-target complexes.[150,173–175,415,416] RNA SELEX exploiting the interaction possibilities of unnatural base pairs could even surpass currently available state-of-the-art aptamer development.

Two similar setups were performed in parallel to address incorporation efficiency of r**TPT3** TP (**23**) and r**NaM** TP (**25**) into RNA during T7 *in vitro* transcription as well as the possibility to reversely transcribe these ribonucleic acids into cDNA adding cognate deoxy triphosphates d**NaM** TP (**24**) or d**TPT3** TP (**22**), respectively.

On the one hand, d**NaM**-containing template **DNA**$^{\mathbf{NaM\text{-}RT}}$ should be transcribed into RNA adding counterpart r**TPT3** TP (**23**) following regular protocols as before giving **RNA**$^{\mathbf{TPT3\text{-}RT}}$ (Figure 150A). Employing a 5′-fluorescently labeled synthetic cDNA primer (**Primer**$^{\mathbf{RT}}$) the obtained RNA should be subjected to rtc reactions with and without the addition of d**NaM** TP (**24**).

On the other hand, **DNA**$^{\mathbf{5SICS\text{-}RT}}$ should give **RNA**$^{\mathbf{NaM\text{-}RT}}$ by transcription with r**NaM** TP (**25**) in an analogous manner (Figure 150B). As mentioned above, the building block of d**TPT3** for solid phase DNA synthesis is not yet available, but the **5SICS** unnatural base is also known to efficiently direct r**NaM** incorporation during transcription.[187,192,205] With the exact same sequence context, the identical **Primer**$^{\mathbf{RT}}$ facilitated enzymatic cDNA synthesis which should be analyzed in presence and absence d**TPT3** TP (**22**) by means of fluorescence scans.

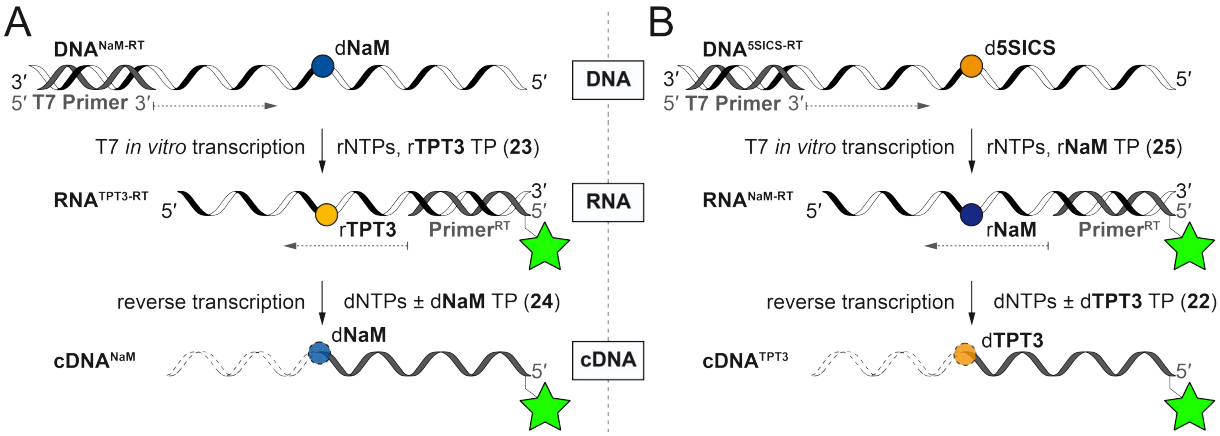


Figure 150. Schematic representation of a transcription-reverse transcription approach involving an expanded genetic alphabet; A. Transcription from **DNA^{NaM-RT}** in the presence of r**TPT3** TP (**23**) and rtc with or without the addition of cognate d**NaM** TP (**24**) B. transcription from **DNA$^{5SICS-RT}$** in the presence of r**NaM** TP (**25**) and rtc with or without the addition of counterpart d**TPT3** TP (**22**).

The 6-carboxyfluorescein (6-FAM) tag at the 5′-end of **Primer^{RT}** enabled facile fluorescent readout which was performed on DPAGE-separated rtcs in multiple independent iterations to ensure properly balanced mean values of the sensitive assay. Figure 151 shows the expected scenarios for reaction outcomes of rtc reactions from unmodified transcripts and putatively UB-modified RNAs with and without the addition of cognate deoxyribose triphosphates corresponding to the unnatural base encoded in the sequence. A native RNA sequence processed by a reverse transcriptase (RT) analyzed on DPAGE would result in detection of the full-length cDNA and likely an excess primer band (Figure 151, Lane 1).

Rtc from a ribonucleic acid containing an unnatural base in the absence of its cognate deoxy triphosphate could result in full skipping or mutagenic read-through thus presenting a similar pattern. However, stalling events at the unnatural base site are also likely to occur; disability of an RT enzyme to bypass the UB could thus result in complete stop leaving a truncated cDNA fragment (Figure 151, Lane 2). Also partial ability to handle the unnatural base is conceived which would give an excess primer band, stalling represented by a truncated cDNA and the full reverse transcript as shown in Figure 151, Lane 3. This reaction outcome would also be observed if the RT is not able to read over the UB, but the RNA does only partially bear the modification.

When an unnatural base-modified RNA such as **RNA^{NaM-RT}** or **RNA$^{5SICS-RT}$** is reversely transcribed in the presence of a counterpart deoxytriphosphate, full processibility would only give full-length cDNA. Also complete deficiency of dUB TP conversion and decent ability to recognize and incorporate d**TPT3** TP (**22**) or d**NaM** TP (**24**) are conceivable.

Therefore, this approach cannot distinguish between skipping/mutagenic read-through and incomplete modification in UB-containing RNA. However, differences comparing rtc experiments with and without added cognate triphosphate reveal the RT's properties when encountering the **TPT3**:**NaM** base pair.

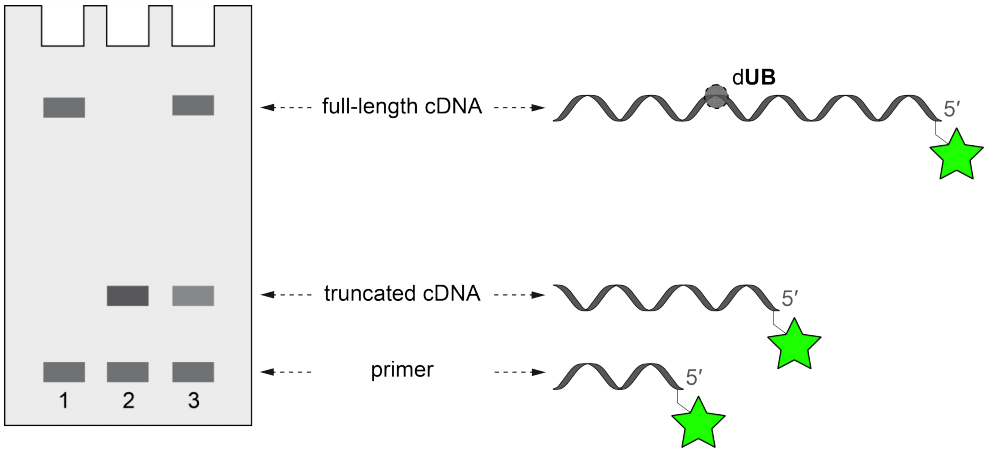


Figure 151. Schematic representation of rtc scenarios with a fluorescent primer analyzed via DPAGE separation; Lane 1: Unmodified RNA; Lane 2: Fully unnatural base-modified RNA in the absence of cognate deoxy triphosphate without RT read-through; Lane 3: Partially unnatural base-modified RNA in the absence of cognate deoxy triphosphate without RT read-through or fully unnatural base-modified RNA in the absence of cognate deoxy triphosphate with read-through.

Experiments on transcription, reverse transcription and optimization of the workflow as well as major parts of data evaluation were carried out during the bachelor thesis of K. Kurscheidt[372] and an internship of E. Hoffmann[371].

A total of five different commercially available reverse transcriptases were tested. In detail, Avian myeloblastosis virus (AMV) RT, moloney murine leukemia virus (MMLV) RT, SuperScript II (SS II) and SuperScript IV (SS IV) RTs as well as the Volcano2G (V2G) polymerase. AMV and MMLV RTs are retroviral enzymes, SS II and SS IV have been genetically engineered from MMLV RT. Retroviruses are known to introduce many mutations enabling to overcome the defense mechanism of infected host organisms.[116,417] Therefore, retrovirally-derived RTs possibly possess the ability to accept and process unnatural bases. V2G polymerase has been developed from *Taq* DNA polymerase and exhibits reverse transcriptase activity.[418] As *Taq* DNA polymerase is able to efficiently replicate the **TPT3**:**NaM** UBP[194] it is a promising candidate to reversely transcribe RNA containing these nucleotides.

Each of these enzymes was multiply tested in a reliable manner on both **RNA**$^{\textbf{TPT3-RT}}$ and **RNA**$^{\textbf{NaM-RT}}$ with and without the addition of the respective deoxyribose triphosphate counterpart. Control reactions on the same sequence without modification (**RNA**$^{\textbf{n-RT}}$) were

executed to rule out structural elements hindering rtc at the UB position thus giving falsified readout information.

Figure 152 shows representative DPAGE analyses of rtc reactions with AMV RT. Facile assignment of the obtained bands was achieved using a custom marker (M) composed of fluorescently labeled DNA oligonucleotides with sequence lengths corresponding to primer, truncated, and full-length cDNA products. **RNA$^{n\text{-}RT}$** served as control template ensuring unhindered propagation of the template.

AMV RT predominantly stalled during rtc from **RNA$^{TPT3\text{-}RT}$**, which is not substantially changed upon addition of counterpart d**NaM** TP (**24**). These findings can be interpreted as near-quantitative and site-specific r**TPT3** modification in the RNA combined with an inability of AMV RT to read over the UB site. The unimproved truncation ratio upon addition of **24** states that the enzyme is not capable to recognize and/or handle the unnatural base pair.

Similar experiments with **RNA$^{NaM\text{-}RT}$** as rtc template resulted in significantly less stopping event at the position of r**NaM**. Therefore, both a low incorporation efficiency of r**NaM** nucleotides during T7 *in vitro* transcription from the d**5SICS**-modified template or a highly increased ability of AMV RT to read over the unnatural base can be conceived. Again, the outcome is not changed significantly when cognate deoxytriphosphate d**TPT3** TP (**22**) is added, neglecting a procession ability.

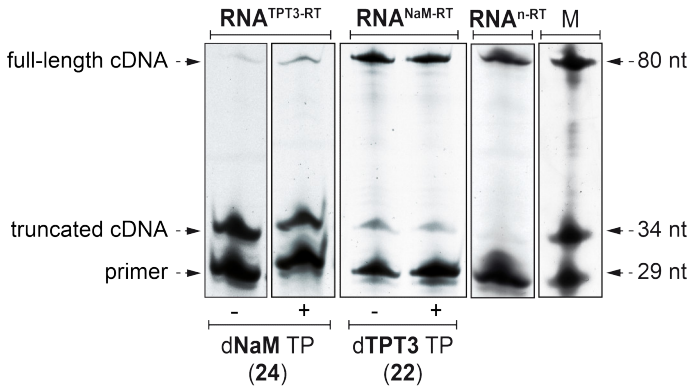


Figure 152. Representative 20 % DPAGE analyses (fluorescence scans) of rtc reactions using AMV RT to reversely transcribe **RNA$^{TPT3\text{-}RT}$** or **RNA$^{NaM\text{-}RT}$** in the presence and absence of their cognate counterpart unnatural deoxy triphosphate as well as control **RNA$^{n\text{-}RT}$**; experiments were performed by K. Kurscheidt[370,372].

In a similar fashion rtcs and analyses thereof were conducted with MMLV, SS II, SS IV, and V2G polymerase. The data obtained were evaluated by band intensities of fluorescent DPAGE readout. Generally, the stop event was used as measurement and represents the ratio of truncation by the sum of truncation and full-length cDNA in percent.

In Figure 153, bar charts represent the stop event during rtc from **RNA$^{TPT3\text{-}RT}$** and **RNA$^{NaM\text{-}RT}$** in the absence of d**NaM** TP (**24**) and d**TPT3** TP (**22**), respectively.

Strikingly, AMV, MMLV, and SS II RT exhibited almost quantitative stalling at the r**TPT3** site. This impressively demonstrates the high incorporation efficiency of r**TPT3** TP (**23**) in template-directed transcription from d**NaM**-modified DNA. Moreover, these polymerases are obviously not able to skip the UB or misincorporate a canonical base opposite r**TPT3**, which makes them excellent candidates to evaluate site-specificity and abundance of r**TPT3**-derived nucleotides in RNA. The SS IV and V2G enzymes stopped to a lower amount, which led to the conclusion that mutagenic read-through is possible in significant amounts.

By contrast, reverse transcription from **RNA**$^{\mathbf{NaM\text{-}RT}}$ consistently gave far less amounts of truncated cDNA. AMV and SS II RT as well as V2G polymerase stopped to ~25 %, MMLV and SS IV RT to an even lower extent. Assuming that AMV RT is not able to read over the unnatural r**NaM** nucleotide, these findings suggest a lower limit of 25 % incorporation efficiency during *in vitro* transcription of the RNA from the d**5SICS**-containing DNA template. This would have far-reaching impact for applications of the full r**TPT3**:r**NaM** UBP in RNA as large shares of modified sequences are essential for many purposes. However, facile mutagenic read-through or skipping of r**NaM** during reverse transcription is also considerable and not distinguishable by this method.

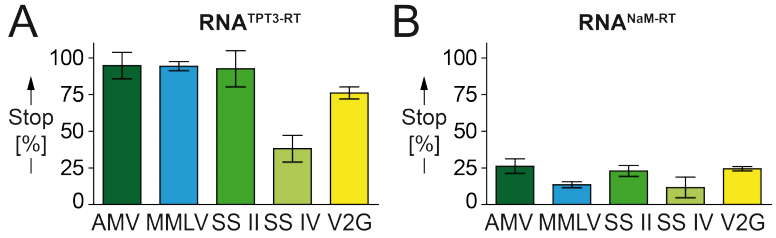


Figure 153. Stop events during rtcs from UB-modified RNAs testing different RTs in the absence of the cognate unnatural base deoxytriphosphate; A. **RNA**$^{\mathbf{TPT3\text{-}RT}}$; B. **RNA**$^{\mathbf{NaM\text{-}RT}}$; data by K. Kurscheidt and E. Hoffmann, each value is the average of at least five independent analyses, error bars are SDs thereof.[370–372]

Reverse transcription experiments employing **RNA**$^{\mathbf{TPT3\text{-}RT}}$ and **RNA**$^{\mathbf{NaM\text{-}RT}}$ with the addition of the particular counterpart deoxynucleoside triphosphate predominantly gave insignificant changes of ±6 % stop event relative to the performance in absence of d**NaM** TP (**24**) or d**TPT3** TP (**22**). Such minor shifts are not considered to be useful for future experiments and generally state inability of the RTs to process the UB TPs.

Notably, SS IV RT responded with an increased stopping event (+35 %) during rtc from **RNA**$^{\mathbf{TPT3\text{-}RT}}$ in the presence of d**NaM** TP (**24**) which indicates that the enzyme can accommodate the unnatural nucleotide in its active site, however it is unable to elongate the cDNA. Furthermore, SS II RT exhibited 12 % less stalling event when d**TPT3** TP (**22**) was added to rtc employing **RNA**$^{\mathbf{NaM\text{-}RT}}$ as template. This suggests certain ability to process the unnatural triphosphate and is so far the only promising candidate to retain UB modifications during reverse transcription from r**NaM**-containing RNA.

Improving the approach, future experiments would benefit from control reactions from synthetically, *i.e.* fully r**NaM**/r**TPT3**-modified RNAs. A prerequisite for this is the synthesis of the respective cyanoethyl phosphoramidites, which have not been reported yet. Analysis of *in vitro* generated reverse transcripts via standard sequencing methods such as Sanger sequencing would be highly desirable. Additionally, preparation of the d**TPT3** building block for solid-phase synthesis of DNA templates suitable for *in vitro* transcription is of considerable interest to investigate whether this UB gives an altered rtc data. Novel engineered RT enzymes that can efficiently accept the dUB TPs as substrates would have great impact on future studies involving UBP technology based on **TPT3**:**NaM**.

Importantly, the presented technique can serve as powerful approach to estimate incorporation efficiency especially of r**TPT3**-derived nucleotides in RNA. For the first time, site-specificity of UB positions in RNA is possible in an unambiguous manner using the reverse transcription approach.

4 Conclusion and outlook

RNA is potentially the most versatile biological macromolecule. For a long time, it was considered only passively transmitting genomic information stored in DNA to facilitate protein synthesis. However, RNA is not limited to encode genetic material, but can also manipulate and regulate cellular pathways as well as perform catalysis. Similar to enzymes, its functional diversity crucially depends on three-dimensional structure development and conformational changes. These are sensitive events that can be easily disturbed by external factors. Studying highly structured non-coding RNAs, many of the available experimental techniques have significant drawbacks. Approaches examining structural features of target sequences are frequently limited by the length of accessible nucleic acids, require laborious and low-yielding multi-step preparation of modified constructs to introduce reporter scaffolds, or involve denaturing conditions. Efficient and straight-forward evaluation of long non-coding RNAs has been hampered using these methods.

Applications using an expanded genetic alphabet can overcome many current shortcomings. With an additional artificial base pair that is efficiently and orthogonally recognized by standard enzymes, the informational density of nucleic acids can be significantly increased. Thus, intrinsic and site-specific insertion of novel entries becomes feasible. The unnatural base pair (UBP) **TPT3**:**NaM** by the Romesberg group is PCR-amplified with natural-like efficiency and fidelity and can even serve to create semi-synthetic organisms with an augmented genetic repertoire.

In this thesis, properties, applications, potential, and limitations of an expanded genetic alphabet have been extensively investigated to enable generation of site-specifically modified and sensitively folded functional non-coding RNA constructs via T7 *in vitro* transcription.

Centrally, development and optimization of a synthetic route towards cyclopropene-functionalized triphosphate r**TPT3^{CP}** TP (**34**) allowed for template-directed incorporation of this novel nucleotide at predefined sites into RNA via expanded genetic alphabet transcription making use of the UBP **TPT3^{CP}**:**NaM** (Figure 154).

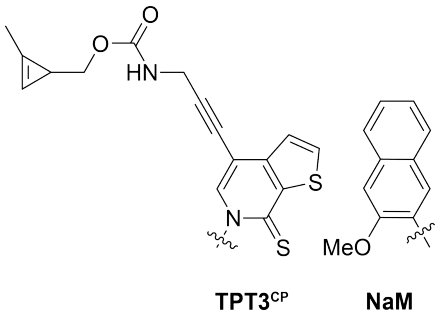


Figure 154. The UBP **TPT3^{CP}**:**NaM**, sugar residues are omitted for clarity.

Several approaches have been established in this work to sequentially facilitate introduction of r**TPT3^{CP}** nucleotides into ribonucleic acid sequences of any length.

Chemically synthesized DNA templates containing d**NaM** nucleotides were employed for *in vitro* transcription in the presence of r**TPT3^{CP}** TP (**34**) to prepare short site-specifically modified RNAs. The highly reactive cyclopropene mini-tags allowed for post-transcriptional attachment of reporter groups such as fluorophores by ultra-mild and bio-orthogonal inverse electron-demand Diels-Alder (IEDDA) cycloaddition reactions with tetrazine derivatives. Outstanding template-directed incorporation efficiency and specificity of r**TPT3^{CP}** nucleotides by T7 RNA polymerase was proved on various levels.

Deoxynucleoside triphosphates of the UBP **TPT3**:**NaM** were synthesized according to literature procedures ensuring stable propagation of the UBP in PCR amplification of modified DNAs. By this, construct assembly beyond restrictions of chemical oligonucleotide synthesis was possible via 6-letter fusion PCR. From these, site-specifically r**TPT3^{CP}**-labeled *glmS* ribozyme sequences were prepared in a modular approach and post-transcriptionally decorated with fluorophores by IEDDA reactions. Pronounced cleavage activity of the ribozymes verified minimal impact on the sensitive secondary structure. Therefore, this method was confirmed to hold enormous potential for the investigation of functional non-coding RNA.

Insertion of the **TPT3**:**NaM** UBP into a plasmid DNA was employed to transcribe a 401 nt long fragment of the Xist RNA site-specifically labeled with an r**TPT3^{CP}** nucleotide. Facile preparation of modified RNA exceeding the limits of solid-phase synthesis could thus be confirmed. These achievements allow for bio-orthogonal attachment of virtually any reporter group to arbitrary targets in prospective non-coding RNA research efforts.

Furthermore, nitroxide-derivatized **TPT3** triphosphate r**TPT3^{NO}** TP (**61**), developed by C. Domnick, was used to site-specifically introduce stable radicals into RNA via *in vitro* transcription. Self-complementary duplexes containing these spin label nucleotides were employed for electron paramagnetic resonance (EPR)-based measurements in cooperation with G. Hagelueken and O. Schiemann. Applicability of the approach was verified with two self-complementary RNA duplexes of different length and revealed a high resolution nanometer-range ruler for inter-spin distance determination in RNA complexes, which was confirmed by excellent agreement with computationally obtained values.

As model system for larger constructs with complex structure, *glmS* ribozyme sequences were labeled with two r**TPT3^{NO}** entries. Distance probability densities were obtained via pulsed electron-electron double resonance (PELDOR) EPR spectroscopy. Supported by molecular dynamics (MD) calculations performed by S. Kath-Schorr, these distances could be faithfully assigned to match the experimental outcome. Several bottlenecks such as

3′-unspecific incorporation of r**TPT3**$^{\mathbf{NO}}$ nucleotides by T7 RNA polymerase were sequentially addressed. Ongoing research on this project is now provided with all precautionary measures necessary for efficient spin labeling and viable distance determination to gather information on structural dynamics of RNA in solution without length limitations.

Furthermore, considerable progress was achieved investigating synthetic accessibility towards a diarylethene-based photoswitchable unnatural nucleotide. Various challenges were encountered and circumvented to identify the optimal route for an efficient preparation of a novel RNA building block with fascinating future implementations. Completion of the proposed synthesis sequence and site-specific introduction into RNA oligonucleotides can serve to enable applications on intriguing research goals such as spatial and temporal control of structural features and thus intra- and intermolecular interactions.

Together with pioneering syntheses of further ribonucleic acid building blocks related to the UBP **TPT3**:**NaM** and development of a fluorescence-based reverse transcription approach that facilitates identification and quantification of site-specific unnatural nucleotide modifications in RNA, the base for a plethora of future applications was set.

The available toolbox for site-specific RNA labeling approaches has been augmented by use of an expanded genetic alphabet by this work. Non-disturbing introduction of functionalized orthogonal nucleotides by standard T7 *in vitro* transcription offers various advantages over previous techniques. High potential for applications in a cellular environment will promote and accelerate progress on functional non-coding RNA investigations. This work will thus help deepening the understanding of currently significantly underexplored RNA species, which may in turn contribute to fundamental research such as unraveling essential aspects of chemical biology and furthermore support development of unprecedented therapeutics.

5 Materials and methods

In the following chapter all equipment, materials, procedures and handling necessary for this study are described in detail, including preceding interdisciplinary methodologies, chemical, and biochemical methods.

5.1 Software and data analysis

Table 4 shows all relevant software applications for data evaluation, processing, and depiction employed in this study.

Table 4. Software applications used in this study.

Software	Version/s	Source
Creative Suite	3–5	*Adobe*
AIDA	4.27.039	*Elysia-Raytest*
ChemDraw	16.0.1.4	*PerkinElmer*
MestReNova	8.0.1	*Mestrelab Research S.L.*
Office	2007	*Microsoft*
Mongo Oligo Mass Calculator	2.08	*https://mods.rna.albany.edu/masspec/Mongo-Oligo*
OligoCalc Oligo Properties Calculator	3.27	*http://biotools.nubic.northwestern.edu/OligoCalc.html*
Origin	8.0.63.988 SR6	*OriginLab Corporation*
PyMol	2007	*DeLano Scientific*
Compass DataAnalysis	4.2	*Bruker*
Prism	5.01	*GraphPad*

5.2 Chemical methods

5.2.1 General chemical methods and equipment

Chemicals and suppliers

All commercial chemicals employed in synthetic procedures were purchased and used without further treatment or purification unless otherwise stated. Suppliers were *abcr* (Karlsruhe, Germany), *Apollo Scientific* (Bredbury, UK), *Applichem* (Darmstadt, Germany), *Berry & Associates* (Dexter, MI, USA), *Carbolution* (St. Ingbert, Germany), *Carbosynth* (Compton, UK), *Carl Roth* (Karlsruhe, Germany), *Chemspace* (Riga, Latvia), *Jena Bioscience* (Jena, Germany), *Merck* including *Sigma Aldrich* (Darmstadt, Germany), *Oxchem* (Wood Dale, IL, USA), *Thermo Fisher Scientific* including *Acros Organics, Alfa Aesar*, and *Honeywell Fluka* (Waltham, MA, USA), *TCI* (Tokyo, Japan), and *VWR* (Radnor, PA, USA).

Schlenk technique

Syntheses referred to as "under inert atmosphere", "under an atmosphere of argon" or similar, were performed using standard Schlenk technique (*i.e.* air- and moisture-free)[419,420] including heat drying of employed glassware under high vacuum and purging with dry argon gas (N46, *Air Liquide*) in a 3× vacuum/argon iteration using a Schenk line. Where particular absence of moisture and/or air was crucial for reaction success, syringes and needles were purged with argon gas three times before use.

General remarks

In synthetic procedures drying of organic layers over Na_2SO_4/$MgSO_4$ implies anhydrous salts. Solvent mixtures mentioned in silica gel column chromatography (stationary phase: *Geduran* Si 60 0.063–0.200 mm silica gel, *Merck*) are given in volume by volume (*v*/*v*). Thin layer chromatography (TLC) was used to track reaction progress and to identify target compound fractions during column chromatographic procedures on *Silica Gel 60* F_{254} TLC plates (*Merck*) or *Silica gel on TLC Al foils* (*Sigma Aldrich*). Visualization was carried out in an UV chamber employing a handheld UV lamp at λ = 254 nm. Notably, **36**, **40**, and **41** did not show significant UV absorbance and had to be visualized with a $KMnO_4$ staining reagent (42 mM $KMnO_4$, 482 mM K_2CO_3, 21 mM NaOH). Alternatively, chemical compounds were stained with *Seebach's* stain (13.7 mM phosphomolybdic acid, 30.1 mM $Ce(SO_4)_2$, 1.1 M H_2SO_4) followed by heating with hot air.

Reaction yields were calculated from the limiting starting material (1.0 eq.) if not stated otherwise. Yields of ≥95 % are considered quantitative. Triethylammonium bicarbonate (TEAB) buffer (1.0 M) for quenching and preparative HPLC purification of triphosphate syntheses was prepared by sparging CO_2 through a 1.0 M mixture of HPLC grade NEt_3 with ultrapure water (ddH_2O, see section 5.3.1) until a pH of approx. 7.5 resulted.

5.2.2 NMR spectroscopy

Nucleic magnetic resonance (NMR) spectra were recorded by institutional core facility at room temperature on an *Avance I* 300 MHz, *Avance I* 400 MHz, *Avance I* 500 MHz, *Avance III* HD 500 MHz, or an *Avance III* HD 700 MHz, all from *Bruker*.

Chemical shifts (δ) are given in ppm, coupling constants (J, [Hz]) were calculated from apparent signal positions in first order approximation.

Solvent residual signals (^{1}H and ^{13}C spectra) were referenced as described in the literature.[421] ^{31}P spectra were referenced to orthophosphoric acid as external standard. The particular values employed in this study can be found in Table 5.

Table 5. Chemical shifts of solvent residual signals used as reference in NMR spectra. [421]

solvent	δ ^{1}H [ppm]	δ ^{13}C [ppm]
$CDCl_3$	7.26	77.16
DMSO-d_6	2.50	39.52
CD_3OD	3.31	49.00
D_2O	4.79	-

In case of overlapping or ambiguous signals, "specific assignment given wherever possible" is stated at the beginning of NMR evaluation results. In these cases, the most likely atom number is given first followed by other(s) that possibly could be the origin of the signal, separated by slash punctuation.

5.2.3 HPLC analysis and purification

High performance liquid chromatography (HPLC) was applied to track reaction progresses, to analyze purity of substances, and to isolate product compounds out of complex mixtures. Furthermore RNA transcripts were purified and analyzed by HPLC, as well as nucleoside digestions of oligonucleotides.

Analytic runs and micro purifications of small molecule samples were performed on an *Agilent 1100*, *1200*, or *Agilent 1260 Infinity II* series system (all from *Agilent Technologies*) equipped with a multi wavelength detector (MWD, *Agilent 1100 and 1200*) or a MWD and a fluorescence detector (*1260 Infinity II*). The reference wavelength (default value: λ = 360 nm) was turned off when analyzing samples containing derivatives of the unnatural nucleobase analog **TPT3** due to strong absorbance properties of the thienopyridone scaffold.

Preparative purification of compounds was achieved using a *Gemini NX-C18 AXIA* (5 µm, 110 Å, 75×30 mm) column on an *Agilent 1200 preparative* system (*Agilent Technologies*).

5.2.4 HPLC-MS analysis

High performance liquid chromatography in connection with mass spectrometry (HPLC-MS) was employed to ensure purity and integrity of target compounds and to track reaction progresses. Analyses were performed on an *HTC esquire* (*Bruker Daltonic*) in combination with an *Agilent 1100* Series HPLC system (*Agilent Technologies*). Generally, a standard run was established to analyze small molecule samples at $c \approx 1$ mg mL^{-1} with a gradient of 5→100 % MeCN (solvent B)/0.1 % formic acid (solvent A) in 20 min at a flow rate of 0.4 mL min^{-1}. Triphosphates were generally submitted to HPLC-MS analysis employing 0.1% (*w*/*v*) ammonium acetate as solvent A. Both were analyzed using a *Zorbax Narrow Bore* (2.1×50 mm, 5 µm) C_{18} column (*Agilent Technologies*). In fewer cases of highly hydrophobic molecules (*e.g.* **105**-**107**) a gradient of 20→100 % MeCN/0.1 % formic acid in 20 min at a flow rate of 0.4 mL min^{-1} was employed using the *Zorbax* column. d**NaM** TP (**24**) was analyzed using a 0→60 % MeCN/0.1 % (*w*/*v*) NH_4OAc gradient in 20 min on the same stationary phase. Solvents, gradients, columns and other settings of significance differing from the general protocol are stated within the correspondent paragraph.

5.2.5 (HR) MS spectrometry

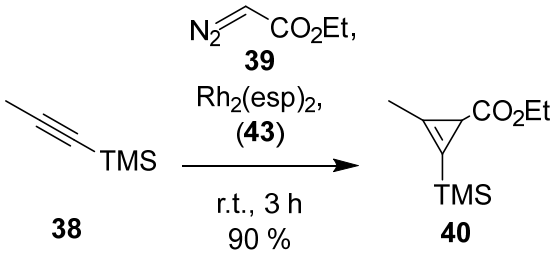
Mass spectrometry (MS) and high resolution mass spectrometry (HR MS) were applied to accurately verify compounds and in particular novel compounds. HR MS spectra (electron spray ionization, ESI) were recorded by institutional core facilities on an *Orbitrap XL* mass spectrometer from *Thermo Fisher Scientific* or on a *micrOTOF-Q* mass spectrometer from *Bruker Daltonic*, occasionally combined with an *Agilent 1200* series HPLC system (*Agilent Technologies*).

5.2.6 Analysis of the IEDDA reaction between **35** and **28** via fluorescence scans

Fluorescence spectra were recorded on a *LS 55* fluorescence spectrometer at room temperature using an Ultra-Micro cuvette (*Hellma*, QS, 45 µL, 3×3 mm). For fluorescence measurements, a solution of cyclopropene-modified nucleoside **35** (60 µM in H_2O/MeOH, 4/1, *v*/*v,* 15 µL) was added to an aqueous solution of tetrazine-fluorophore conjugate **28** (20 µM, 15 µL) in a micro cuvette and directly subjected to iterative fluorescence scans (λ_{ex} = 495 nm, λ_{em} = 450–600 nm, scan rate = 120 nm s^{-1}, slit = 5, measurement start every 83 s, 15 scans). As negative control an equivalent mixture tetrazine-fluorophore conjugate **28** was incubated with iodinated nucleoside **51** for 1 h at room temperature and fluorescence scanned in a similar fashion in triplicate.

5.2.7 Chemical syntheses

5.2.7.1 Synthesis of 2-methyl-3-trimethylsilylcycloprop-2-enecarboxylic acid ethyl ester (40)



Experimental procedure:

The reaction was carried out according to a general cyclopropenation method published by Petiniot *et al.*[299]

Under an atmosphere of argon ethyl diazoacetate (**39**, ≤ 87 % solution in CH_2Cl_2, 1.56 mL, 13.0 mmol, 1.0 eq.) was added slowly over a course of 3.0 h to a suspension of $Rh_2(esp)_2$ (**43**, 100 mg, 0.13 mmol, 1.0 mol %) in trimethylsilyl propyne (**38**, 4.80 mL, 32.5 mmol, 3.0 eq.) at room temperature. After the addition was complete, the resulting mixture was stirred for 3 h at room temperature and concentrated under reduced pressure (100 mbar, 40 °C due to volatility of the product). Purification by column chromatography (EtOAc/Cy, 1/9, R_f = 0.60) yielded title compound **40** (2.33 g, 11.7 mmol, 90 %) as faint yellow oil.

Analytics:

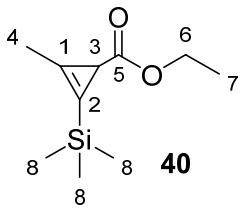
Chemical Formula: $C_{10}H_{18}O_2Si$.

Molecular Weight [g mol^{-1}]: 198.33.

^{1}H-NMR ($CDCl_3$, 300 MHz): δ [ppm] = 4.13–4.05 (m; 2H; H-6), 2.18 (s; 3H; H-4), 1.97 (s; 1H; H-3), 1.22 (t; $^3J_{H\text{-}7/H\text{-}6}$ = 7.1 Hz; 3H; H-7), 0.18 (s; 9H; H-8).

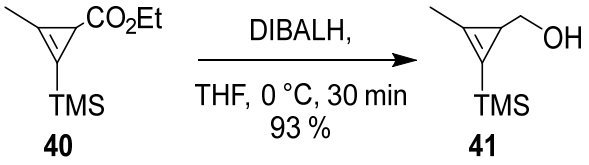
^{13}C-NMR ($CDCl_3$, 75 MHz): δ [ppm] = 177.2 (C-5), 122.7 (C-1), 104.3 (C-2), 59.9 (C-6), 21.4 (C-3), 14.6 (C-4), 12.0 (C-7), -1.4 (C-8).

MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{10}H_{18}O_2SiNa^+$): 221.1; found: *m*/*z* = 221.1.



40

5.2.7.2 Synthesis of (2-methyl-3-trimethylsilylcycloprop-2-enyl)methanol (41)



The following protocol was adapted from Yang *et al.*[239]

Under an atmosphere of argon ester **40** (283 mg, 1.4 mmol, 1.0 eq) was added dropwise to a stirred solution of DIBALH in THF (*c* = 1.0 M, 3.1 mL, 3.1 mmol, 2.2 eq.) at 0 °C. Stirring was continued for 30 min at 0 °C. The reaction was then quenched by the addition of saturated aqueous sodium potassium tartrate solution (3 mL). The mixture was allowed to warm to room temperature and stirred for several minutes until a colorless gel formed. The aqueous gel was shredded by vigorous stirring and subsequently extracted with Et_2O. The combined organic layers were dried over Na_2SO_4. Removal of the solvent under reduced pressure gave alcohol **41** (207 mg, 1.3 mmol, 93 %) as a colorless oil which was submitted to the next reaction without further purification.

Analytics:

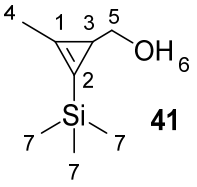
Chemical Formula: $C_8H_{16}O_2Si$.

Molecular Weight [g mol^{-1}]: 156.10.

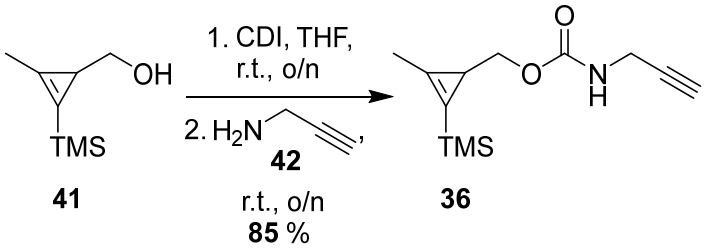
^{1}H-NMR ($CDCl_3$, 300 MHz): δ [ppm] = 3.48 (d; $^3J_{H\text{-}5/H\text{-}3}$ = 4.6 Hz; 2H; H-5), 2.21 (s; 3H; H-4), 1.56 (t; $^3J_{H\text{-}3/H\text{-}5}$ = 4.6 Hz; 1H; H-3), 0.16 (s; 9H; H-7).

^{13}C-NMR ($CDCl_3$, 75 MHz): δ [ppm] = 135.6 (C-1), 111.4 (C-2), 69.4 (C-5), 22.3 (C-3), 13.6 (C-4), -1.0 (C-7).

MS (ESI$^+$): calculated for $[2M+Na]^+$ ($C_{16}H_{32}O_2Si_2Na^+$): 335.2; found: *m*/*z* = 335.2.



5.2.7.3 Synthesis of (2-methyl-3-(trimethylsilyl)cycloprop-2-en-1-yl)methyl prop-2-yn-1-yl carbamate (36)



Experimental procedure:

Reaction conditions for carbamate formation were adapted from Yang *et al.*[239]

Under argon atmosphere CDI (1.02 g, 6.3 mmol, 1.1 eq.) was dissolved in dry THF (15 mL) and alcohol **41** (900 mg, 5.8 mmol, 1.0 eq.) was added dropwise. After stirring overnight at room temperature, propargylamine (**42**, 0.74 mL, 11.6 mmol, 2.0 eq.) was added and the reaction was agitated overnight. The mixture was diluted with water (20 mL) and the aqueous layer was extracted with Et_2O (3×25 mL). The combined organic layers were dried over Na_2SO_4, and the solvent was removed *in vacuo* to give the crude product. Purification by column chromatography (Cy/EtOAc, 5/3, R_f = 0.85) yielded **36** (1.17 g, 4.9 mmol, 85 %) as a colorless oil.

Analytics:

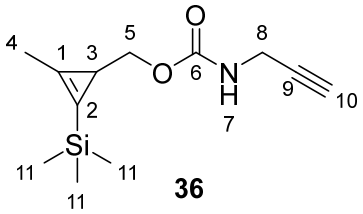
Chemical Formula: $C_{12}H_{19}NO_2Si$.

Molecular Weight [g mol^{-1}]: 237.37.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 4.84 (bs; 1H; H-7), 4.00–3.86 (m; 4H; H-5; H-8), 2.22 (t; $^4J_{H\text{-}10/H\text{-}8}$ = 2.5 Hz; 1H; H-10), 2.18 (s; 3H; H-4), 1.53 (t; $^3J_{H\text{-}3/H\text{-}5}$ = 5.3 Hz; 1H; H-3), 0.14 (s; 9H; H-11).

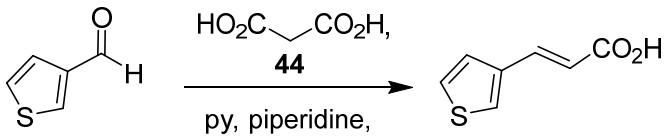
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 156.6 (C-6), 134.5 (C-1), 111.1 (C-2), 80.1 (C-9), 74.1 (C-5), 71.5 (C-10), 30.9 (C-8), 18.6 (C-3), 13.3 (C-4), -1.1 (C-11).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{12}H_{19}NO_2SiNa^+$): 260.1077; found: *m*/*z* = 260.1075.



36

5.2.7.4 Synthesis of (2*E*)-3-(3-thienyl)-2-propenoic acid (19)



18 **19**

Experimental procedure:

Synthesis of acrylic acid **19** was achieved using a modified method by New *et al.*[306]

Carboxaldehyde **18** (10.0 g, 89.2 mmol, 1.0 eq.) and malonic acid (**44**, 9.28 g, 89.2 mmol, 1.0 eq.) were dissolved in pyridine (100 mL), then piperidine (4.4 mL, 44.6 mmol, 0.5 eq.) was added. The solution was refluxed for 2 h, cooled to ambient temperatures, and poured into concentrated $HCl_{aq.}$ (150 mL) under ice cooling. The suspension was extracted with

EtOAc (3×200 mL) and the combined organic layers were dried over $MgSO_4$. Evaporation of solvents yielded acrylic acid **19** (13.1 g, 84.8 mmol, 95 %) as beige solid.

Analytics:

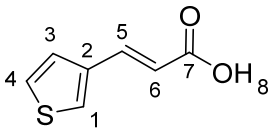
Chemical Formula: $C_7H_6O_2S$.

Molecular Weight [g mol^{-1}]: 154.19.

^{1}H-NMR ($CDCl_3$, 300 MHz): δ [ppm] = 7.76 (d; $^3J_{H\text{-}5/H\text{-}6}$ = 15.9 Hz; 1H; H-5), 7.54 (dd; $^4J_{H\text{-}1/H\text{-}3}$ = 2.9 Hz; $^4J_{H\text{-}1/H\text{-}4}$ = 0.9 Hz; 1H; H-1), 7.34 (ddd; $^3J_{H\text{-}3/H\text{-}4}$ = 5.1 Hz; $^4J_{H\text{-}3/H\text{-}1}$ = 2.9 Hz; $^4J_{H\text{-}3/H\text{-}5}$ = 0.4 Hz; 1H; H-3), 7.31 (dd; $^3J_{H\text{-}4/H\text{-}3}$ = 5.1 Hz; $^4J_{H\text{-}4/H\text{-}1}$ = 0.9 Hz; 1H; H-4), 6.26 (d; $^3J_{H\text{-}6/H\text{-}5}$ = 15.9 Hz; 1H; H-6).

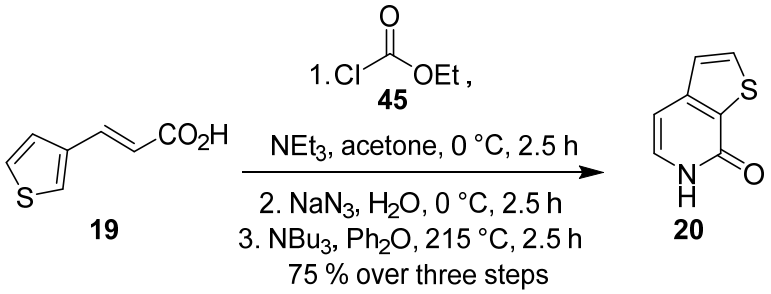
^{13}C-NMR ($CDCl_3$, 75 MHz): δ [ppm] = 172.9 (C-7), 140.6 (C-5), 137.4 (C-2), 129.2 (4/1/3), 127.3 (C-1/3/4), 125.4 (C-3/1/4), 117.1 (C-6).

MS (ESI$^-$): calculated for $[M\text{-}H]^-$ ($C_7H_5O_2S^-$): 153.0; found: *m*/*z* = 153.0.



19

5.2.7.5 Synthesis of 6*H*-thieno[2,3-*c*]pyridin-7-one (20)



Experimental procedure:

Pyridone **20** was synthesized according to the description within an US patent of Ewing *et al.*[307]

3-Thiophen-3-ylacrylic acid (**19**, 12.0 g, 77.8 mmol, 1.0 eq.) and NEt_3 (11.9 mL, 85.6 mmol, 1.1 eq.) were dissolved in acetone (100 mL) and cooled in an ice bath. Ethyl chloroformate (**45**, 8.2 mL, 85.6 mmol, 1.1 eq.) was added dropwise and the mixture was stirred for 1 h at 0 °C. Subsequently, a solution of NaN_3 (7.6 g, 116.7 mmol, 1.5 eq.) in water (20 mL) was added. After continued stirring at 0 °C for 1 h, the reaction mixture was poured into water (500 mL). The resulting precipitate was collected by filtration, washed with water and dried *in vacuo*. Dissolved in Ph_2O (20 mL), the solid was added to NBu_3 (20.3 mL, 85.6 mmol, 1.1 eq) in Ph_2O (80 mL) at 2305 °C. The mixture was stirred at 230 °C for 3.5 h, allowed to cool to ambient temperatures and poured into cold *n*-hexane (500 mL). The occurring solid was

collected by filtration and washed with *n*-hexane. Drying under high vacuum gave **20** (8.80 g, 58.1 mmol, 75 %, Lit.: 72 %[307]) as a light brown solid.

Analytics:

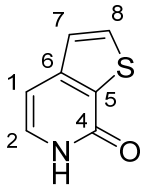
Chemical Formula: C_7H_5NOS.

Molecular Weight [g mol^{-1}]: 151.19.

^{1}H-NMR (DMSO-*d*$_6$, 400 MHz): δ [ppm] = 11.47 (bs; 1H; H-3), 8.02 (d; $^3J_{H\text{-}8/H\text{-}7}$ = 5.2 Hz; 1H; H-8), 7.37 (d; $^3J_{H\text{-}7/H\text{-}8}$ = 5.2 Hz; 1H; H-7), 7.26 (d; $^3J_{H\text{-}2/H\text{-}1}$ = 6.9 Hz; 1H; H-2), 6.71 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 6.9 Hz; 1H; H-1).

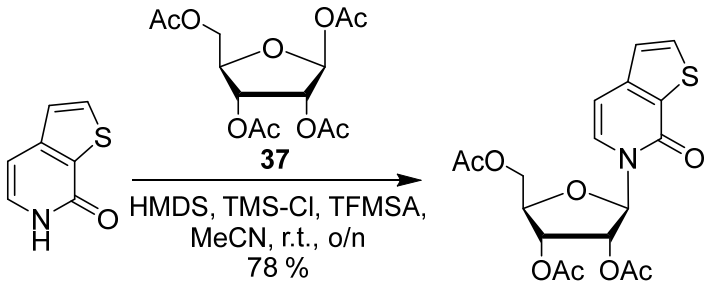
^{13}C-NMR (DMSO-*d*$_6$, 101 MHz): δ [ppm] = 158.5 (C-4), 146.2 (C-6), 133.7 (C-8), 130.1 (C-2), 129.1 (C-5), 124.9 (C-7), 102.1 (C-1).

MS (ESI$^+$): calculated for $[M+H]^+$ ($C_7H_6NOS^+$): 152.0; found: *m*/*z* = 152.0.



20

5.2.7.6 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)6*H*-thieno-[2,3-*c*]pyridin-7-one (48)



20 **48**

Experimental procedure:

The following reaction was performed according to a slightly altered description by Vorbrüggen and Bennua.[317]

Pyridone derivative **20** (5.00 g, 33.1 mmol, 1.0 eq.) was dissolved in dry MeCN (200 mL) under an atmosphere of argon. HMDS (6.9 mL, 33.1 mmol, 1.0 eq.) and TMS-Cl (5.1 mL, 40.0 mmol, 1.2 eq.) were added consecutively, the mixture was stirred for 1 h at room temperature. After the addition of peracylated β-D-ribofuranose (**37**, 12.7 g, 40.0 mmol, 1.2 eq.) and TFMSA (3.5 mL, 40.0 mmol, 1.2 eq.) the reaction mixture was stirred at room temperature overnight. Neutralization was performed with saturated aqueous $NaHCO_3$ solution (200 mL) by vigorous stirring for 15 min. Extraction of the aqueous layer with CH_2Cl_2

(3×200 mL), drying over anhydrous Na_2SO_4 and removal of the solvents under reduced pressure gave the crude product which was purified by column chromatography (EtOAc/Cy, 7/4, R_f = 0.60) to yield **48** (10.6 g, 25.8 mmol, 78 %) as a light brown oil of very high viscosity.

Analytics (specific assignment given wherever possible):

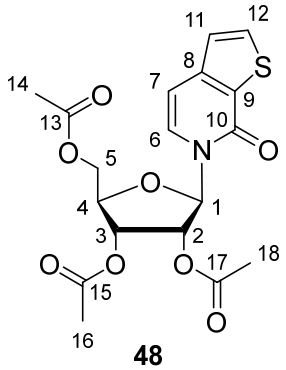
Chemical Formula: $C_{18}H_{19}NO_8S$.

Molecular Weight [g mol^{-1}]: 409.41.

^{1}H-NMR (CDCl$_3$, 400 MHz): δ [ppm] = 7.71 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.1 Hz; 1H; H-12), 7.36 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.4 Hz; 1H; H-6), 7.19 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.1 Hz; 1H; H-11); 6.69 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.4 Hz; 1H; H-7), 6.46 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 4.5 Hz; 1H; H-1), 5.47 (m; 1H; H-2/3), 5.41 (m; 1H; H-3/2), 4.45–4.34 (m; 3H; H-4, H-5), 2.15 (s; 3H; H-18), 2.10 (s; 3H; H-16/14), 2.09 (s; 3H; H-14/16).

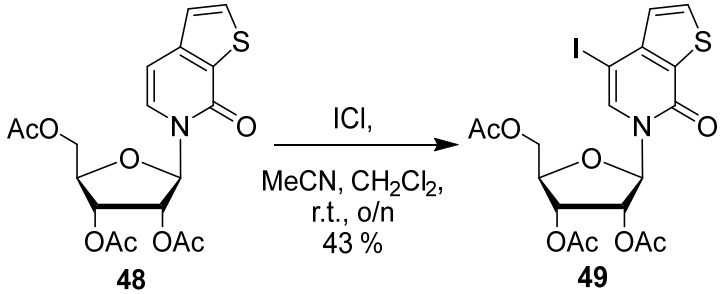
^{13}C-NMR (CDCl$_3$, 101 MHz): δ [ppm] = 170.41 (C-13), 169.73 (C-15/17), 169.66 (C-17/15), 158.30 (C-10), 145.27 (C-8), 134.39 (C-12), 130.27 (C-9), 127.82 (C-6), 124.39 (C-11), 103.87 (C-7), 87.88 (C-1), 79.59 (C-4), 74.03 (C-3), 70.13 (C-2), 63.21 (C-5), 20.91 (C-14), 20.62 (C-16/18), 20.59 (C-18/16).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{18}H_{19}NO_8SNa^+$): 432.0724; found: *m*/*z* = 432.0733.



48

5.2.7.7 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)-6*H*-thieno-[2,3-*c*]-4-iodopyridin-7-one (49)



48 **49**

Experimental procedure:

Iodination was achieved applying a protocol by Li *et al*.[194]

Under an atmosphere of argon compound **48** (2.30 g, 5.6 mmol, 1.0 eq.) was dissolved in dry MeCN (50 mL) and cooled in an ice bath. Iodine monochloride (1.0 M soln. in CH_2Cl_2,

8.6 mL, 8.6 mmol, 1.5 eq.) was added and the mixture was stirred overnight at room temperature under exclusion of light. The reaction was quenched by the addition of saturated aqueous $NaHCO_3$ and $Na_2S_2O_3$ solution (50 mL each). The aqueous layer was extracted with CH_2Cl_2 (3×100 mL), dried over $MgSO_4$ and concentrated *in vacuo*. Purification via column chromatography (EtOAc/Cy, 6/4, R_f = 0.60) yielded **49** (1.29 g, 2.4 mmol, 43 %) as white foam.

Analytics (specific assignment given wherever possible):

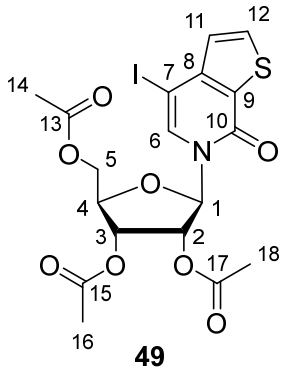
Chemical Formula: $C_{18}H_{18}INO_8S$.

Molecular Weight [g mol^{-1}]: 535.31.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 7.75 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.2 Hz; 1H; H-12), 7.69 (s; 1H; H-6), 7.20 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.2 Hz; 1H; H-11), 6.45 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 4.5 Hz; 1H; H-1), 5.43-5.37 (m; 2H; H-2, H-3), 4.45–4.36 (m; 3H; H-4, H-5), 2.26 (s; 3H; H-14), 2.10 (s; 3H; H-16/18), 2.09 (s; 3H; H-18/16).

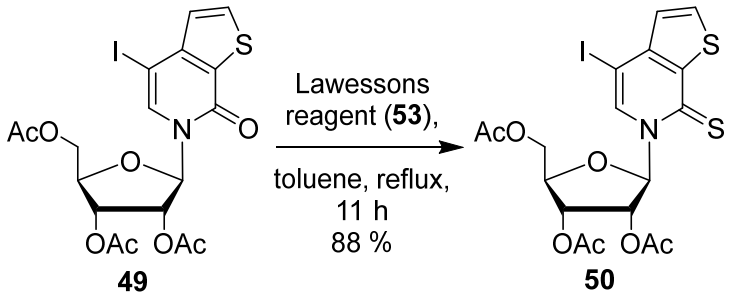
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 170.41 (C-13), 169.70 (C-15/17), 169.55 (C-17/15), 157.39 (C-10), 147.06 (C-8), 133.93 (C-12), 132.81 (C-6), 129.21 (C-9), 128.27 (C-11), 87.29 (C-1), 79.90 (C-4), 74.23 (C-2), 69.93 (C-3), 64.58 (C-7), 62.89 (C-5), 21.29 (C-14), 20.60 (C-16/18), 20.55 (C-18/16).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{18}H_{18}NO_8ISNa^+$): 557.9690; found: *m*/*z* = 557.9703.



49

5.2.7.8 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)6*H*-thieno-[2,3-*c*]-4-iodopyridine-7-thione (50)



49 **50**

Experimental procedure:

Thionation was achieved by an adapted procedure published by Li *et al*.[194]

In an argon charged system Lawesson's reagent (**53**, 2.98 g, 7.4 mmol, 1.5 eq.) was added to a solution of **49** (2.63 g, 4.9 mmol, 1.0 eq.) in toluene (70 mL). The mixture was refluxed for 11 h, cooled to room temperature, filtrated, and evaporated onto silica. Purification by column chromatography (Cy/EtOAc, 7/4, R_f = 0.43) yielded **50** (2.36 g, 4.3 mmol, 88 %) as bright yellow solid.

Analytics (specific assignment given wherever possible):

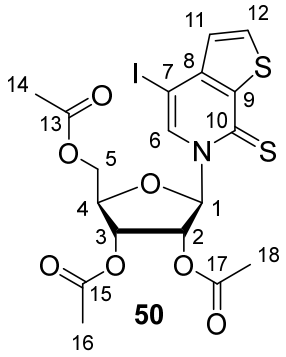
Chemical Formula: $C_{18}H_{18}INO_7S_2$.

Molecular Weight [g mol^{-1}]: 551.37.

^{1}H-NMR (CDCl$_3$, 400 MHz): δ [ppm] = 8.25 (s; 1H; H-6), 7.85 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.3 Hz; 1H; H-12), 7.28 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.3 Hz; 1H; H-11), 7.18 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 2.5 Hz; 1H; H-1), 5.56 (dd; $^3J_{H\text{-}2/H\text{-}1}$ = 2.5 Hz; $^3J_{H\text{-}2/H\text{-}3}$ = 5.5 Hz; 1H; H-2), 5.29 (dd; $^3J_{H\text{-}3/H\text{-}4}$ = 7.1 Hz; $^3J_{H\text{-}3/H\text{-}2}$ = 5.5 Hz; 1H; H-3), 4.54–4.46 (m; 3H; H-4; H-5), 2.32 (s; 1H; H-14), 2.16 (s; 3H; H-16/18), 2.07 (s; 3H; H-18/16).

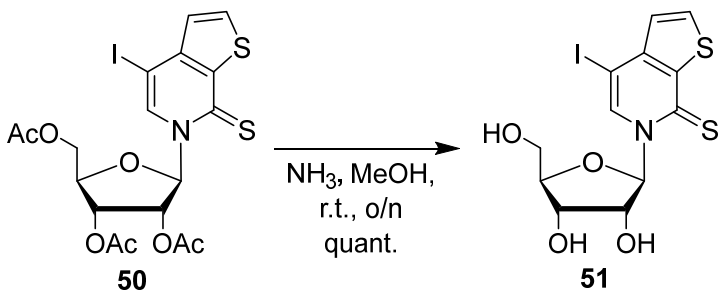
^{13}C-NMR (CDCl$_3$, 101 MHz): δ [ppm] = 173.75 (C-10), 170.48 (C-13), 169.60 (C-15/17), 169.22 (C-17/15), 144.45 (C-9/8), 141.18 (C-8/9), 137.94 (C-12), 134.57 (C-6), 128.30 (C-11), 92.00 (C-1), 79.65 (C-4), 74.61 (C-3), 72.03 (C-7), 68.29 (C-2), 61.70 (C-5), 21.57 (C-14), 20.60 (C-16/18), 20.57 (C-18/16).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{18}H_{18}NO_7IS_2Na^+$): 573.9462; found: *m*/*z* = 573.9471.



50

5.2.7.9 Synthesis of 1′-(β-D-ribofuranosyl)6*H*-thieno-[2,3-*c*]-4-iodo-pyridine-7-thione (51)



50 **51**

Experimental procedure:

Acyl deprotection was carried out adapting a description by Antonini *et al*.[314]

To **50** (667 mg, 1.2 mmol, 1.0 eq.) was added a solution of NH_3 in MeOH (*c* = 7.0 M, 20 mL, 140 mmol, 117 eq.) at room temperature under argon. The resulting solution was set aside overnight. Volatiles were evaporated and the residue was purified by column chromatography (CH_2Cl_2/MeOH, 95/5→90/10, R_f = 0.29 for 95/5) to give **51** (493 mg,1.2 mmol, quant.) as a light yellow solid.

Analytics:

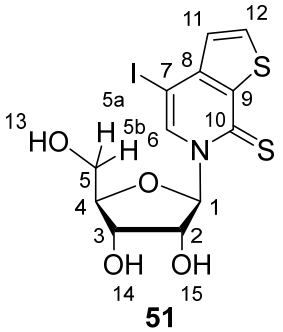
Chemical Formula: $C_{12}H_{12}INO_4S_2$.

Molecular Weight [g mol^{-1}]: 425.26.

^{1}H-NMR (DMSO-*d*$_6$, 400 MHz): δ [ppm] = 9.01 (s; 1H; H-6), 8.25 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.4 Hz; 1H; H-12), 7.33 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.4 Hz; 1H; H-11), 6.81 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 1.2 Hz; 1H; H-1), 5.54 (d; $^3J_{H\text{-}15/H\text{-}2}$ = 4.6 Hz; 1H; H-15), 5.49 (t; 1H; $^3J_{H\text{-}13/H\text{-}5}$ = 4.4 Hz; H-13), 5.07 (d; $^3J_{H\text{-}14/H\text{-}3}$ = 5.5 Hz; 1H; H-14), 4.13–4.03 (m; 3H; H-2, H-3, H-4), 3.89 (ddd; $^2J_{H\text{-}5a/H\text{-}5b}$ = 12.0 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 4.5 Hz; $^4J_{H\text{-}5a/H\text{-}3}$ = 2.0 Hz; 1H; H-5a/5b), 3.68 (ddd; $^2J_{H\text{-}5b/H\text{-}5a}$ = 12.0 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 4.0 Hz; $^4J_{H\text{-}5a/H\text{-}3}$ = 1.3 Hz; 1H; H-5b/5a).

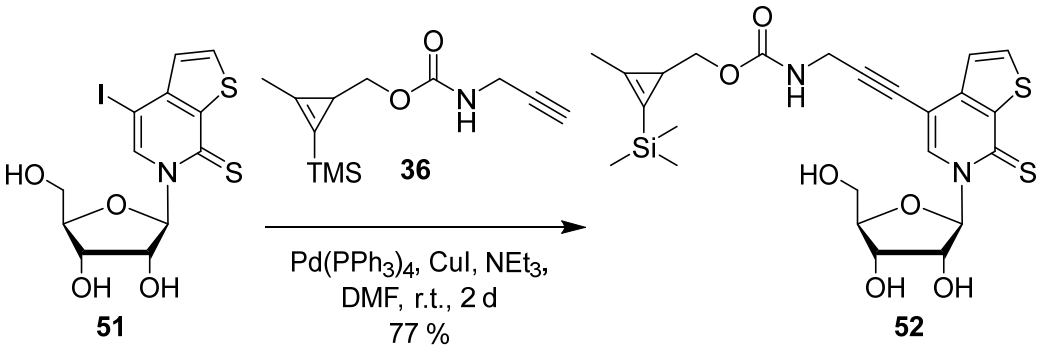
^{13}C-NMR (DMSO-*d*$_6$, 101 MHz): δ [ppm] = 171.82 (C-10), 142.94 (C-9), 141.20 (C-8), 139.25 (C-12), 136.68 (C-6), 127.78 (C-11), 94.33 (C-1), 84.23 (C-2), 75.54 (C-3), 72.52 (C-7), 67.68 (C-4), 58.68 (C-5).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{12}H_{12}NO_4IS_2Na^+$): 447.9145; found: *m*/*z* = 447.9139.



51

5.2.7.10 Synthesis of (2-methyl-3-(trimethylsilyl)cycloprop-2-en-1-yl)methyl (β-D-ribofuranos-1′-yl)-7-thioxo-6,7-dihydrothieno[2,3-*c*]pyridin-4-yl)prop-2-yn-1-yl)carbamate (52)



Experimental procedure:

Reaction conditions for Sonogashira coupling were adapted from Li *et al.*[194]

Under an atmosphere of argon nucleoside **51** (500 mg, 1.18 mmol, 1.0 eq.) and copper(I) iodide (158 mg, 0.83 mmol, 0.7 eq.) were dissolved in dry DMF (40 mL) and the resulting solution was degassed with a stream of argon. Previously degassed NEt_3 (0.50 mL, 3.54 mmol, 3.0 eq.) and a degassed solution of carbamate **36** (335 mg, 1.41 mmol, 1.2 eq.) in dry DMF (10 mL) were added subsequently. After the addition of $Pd(PPh_3)_4$ (139 mg, 0.12 mmol, 0.1 eq.), the mixture was stirred overnight at room temperature under exclusion of light. Another three spatula tips of $Pd(PPh_3)_4$ were added and the reaction was agitated overnight again. The solvent was removed *in vacuo* and the residue was purified by silica gel chromatography (CH_2Cl_2/MeOH, 95/5, R_f = 0.44). Nucleoside **52** (485 mg, 0.91 mmol, 77 %) was obtained as dark yellow solid.

Analytics:

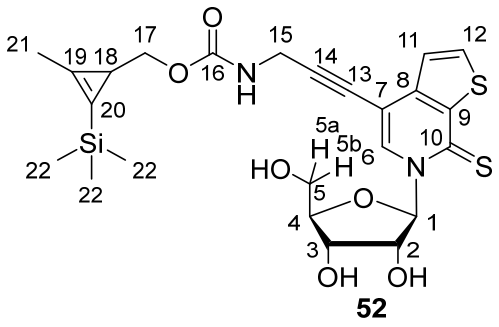
Chemical Formula: $C_{24}H_{30}N_2O_6S_2Si$.

Molecular Weight [g mol^{-1}]: 534.72.

^{1}H-NMR (CD_3OD, 400 MHz): δ [ppm] = 8.78 (s; 1H; H-6), 7.99 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.3 Hz; 1H; H-12), 7.44 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.3 Hz; 1H; H-11), 6.94 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 1.6 Hz; 1H; H-1), 4.26 (dd; $^3J_{H\text{-}2/H\text{-}3}$ = 4.6 Hz; $^3J_{H\text{-}2/H\text{-}1}$ = 1.6 Hz; 1H; H-2), 4.24-4.14 (m; 4H; H-3, H-4, H-15), 4.08 (dd; $^3J_{H\text{-}5a/H\text{-}5b}$ = 12.5 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 2.3 Hz; 1H; H-5a), 3.92 (m; 2H; H-17), 3.88 (dd; $^3J_{H\text{-}5b/H\text{-}5a}$ = 12.5 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 2.3 Hz; 1H; H-5b), 2.19 (s; 3H; H-21), 1.57 (t; $^3J_{H\text{-}18/H\text{-}17}$ = 5.4 Hz; 1H; H-18), 0.15 (s; 9H; H-22).

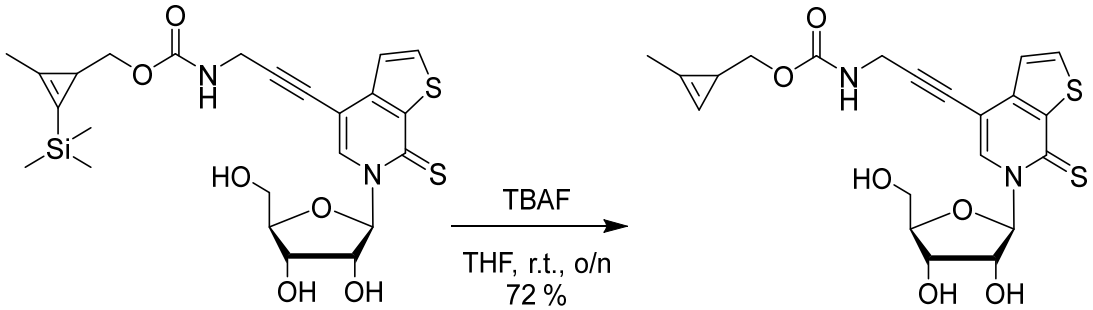
^{13}C-NMR (CD_3OD, 101 MHz): δ [ppm] = 174.1 (C-10), 159.1 (C-16), 145.7 (C-9), 140.5 (C-8), 139.2 (C-12), 136.0 (C-19), 135.5 (C-6), 125.0 (C-11), 112.2 (C-20), 105.9 (C-13), 96.4 (C-1), 92.1 (C-14), 85.6 (C-3), 77.2 (C-2), 74.7 (C-17), 69.2 (C-4), 60.7 (C-5), 32.0 (C-15), 19.9 (C-18), 13.2 (C-21), -1.2 (C-22).

HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{24}H_{31}N_2O_6S_2Si^+$): 535.1387; found: *m*/*z* = 535.1376.



52

5.2.7.11 Synthesis of (2-methylcycloprop-2-en-1-yl)methyl (((β-D-ribofuranos-1′-yl)-7-thioxo-6,7-dihydrothieno[2,3-*c*]pyridin-4-yl)prop-2-yn-1-yl)carbamate (35, rTPT3^{CP})



52 **35**

Experimental procedure:

TMS deprotection was achieved adapting a description by Yang *et al.*[239]

Nucleoside **52** (190 mg, 0.36 mmol, 1.0 eq.) was dissolved in dry THF (10 mL). Subsequently, tributylammonium fluoride trihydrate (150 mg, 0.48 mmol, 1.3 eq.) was added and the mixture was stirred at room temperature overnight. The solvent was removed *in vacuo* and the crude product was purified *via* silica gel chromatography (CH_2Cl_2/MeOH, 95/5, R_f = 0.37). Methyl cyclopropene nucleoside r**TPT3^{CP}** (**35**) was obtained as yellow foam (119 mg, 0.26 mmol, 72 %).

Analytics:

Chemical Formula: $C_{21}H_{22}N_2O_6S_2$.

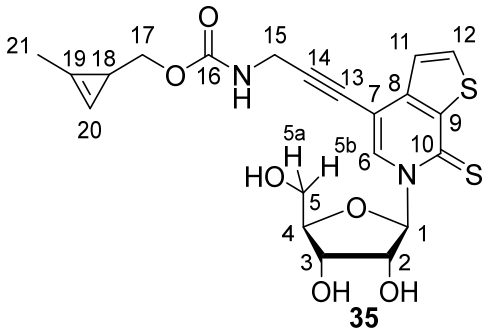
Molecular Weight [g mol^{-1}]: 462.54.

^{1}H-NMR (CD_3OD, 400 MHz): δ [ppm] = 8.80 (s; 1H; H-6), 8.00 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.3 Hz; 1H; H-12), 7.46 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.3 Hz; 1H; H-11), 6.95 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 1.5 Hz; 1H; H-1), 6.65 (s; 1H; H-20), 4.25 (dd; $^3J_{H\text{-}2/H\text{-}3}$ = 4.6 Hz; $^3J_{H\text{-}2/H\text{-}1}$ = 1.6 Hz; 1H; H-2), 4.23–4.16 (m; 4H; H-3; H-4; H-15), 4.08 (dd; $^3J_{H\text{-}5a/H\text{-}5b}$ = 12.5 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 2.3 Hz; 1H; H-5a), 4.04–3.86 (m; 3H; H-17; H-5b), 2.12 (d; $^3J_{H\text{-}21/H\text{-}20}$ = 1.2 Hz; 3H; H-21), 1.64 (t; $^3J_{H\text{-}18/H\text{-}17}$ = 4.1 Hz; 1H; H-18).

^{13}C-NMR (CD_3OD, 101 MHz): δ [ppm] = 174.2 (C-10), 159.1 (C-16), 145.7 (C-9), 140.6 (C-8), 139.3 (C-12), 135.6 (C-6), 125.0 (C-11), 122.2 (C-19), 105.9 (C-13), 102.8 (C-20), 96.4

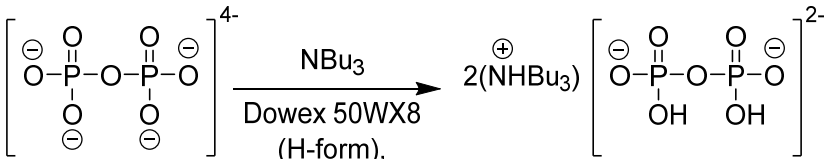
(C-1), 92.1 (C-14), 85.7 (C-3), 77.3 (C-2), 73.6 (C-17), 69.2 (C-4), 60.7 (C-5), 31.9 (C-15), 18.3 (C-18), 11.6 (C-21).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{21}H_{22}N_2O_6S_2Na^+$): 485.0811; found: *m*/*z* = 485.0810.



35

5.2.7.12 Synthesis of bis(tri-*n*-butylammonium) pyrophosphate (127)



128 **127**

Experimental procedure:

Generation of pyrophosphate species **127** was adapted from Moffat[334,422].

Dowex 50WX8 ion exchange resin (100 g) was stirred in $HCl_{aq.}$ (200 mL, 10 %, *v*/*v*) for 30 min at room temperature. Water (approx. 100 mL) was passed through the resin in a fritted glass column until the eluate appeared to be neutral (pH = 7). A solution of tetrasodium pyrophosphate decahydrate (**128**, 10 g, 22.4 mmol, 1.0 eq) in water (150 mL) was eluted into a solution of NBu_3 (11.8 mL, 49.6 mmol, 2.2 eq.) and EtOH (150 mL). The resin was flushed with water (approx. 200 mL) and the total effluent collected. Removal of solvents was carried out under reduced pressure and the residue was freeze-dried over four days. Coevaporation with anhydrous pyridine (3×30 mL) afforded bis(tri-*n*-butylammonium) pyrophosphate (**127**, 10.02 g, 18.3 mmol, 82 %) as white solid.

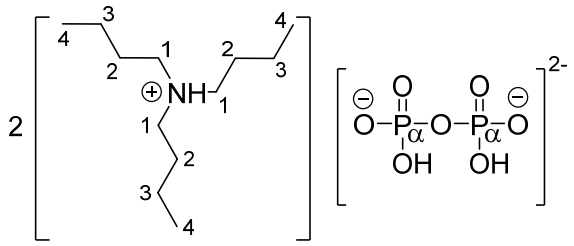
Analytics:

Chemical Formula: $C_{24}H_{58}N_2O_7P_2$.

Molecular Weight [g mol^{-1}]: 548.37.

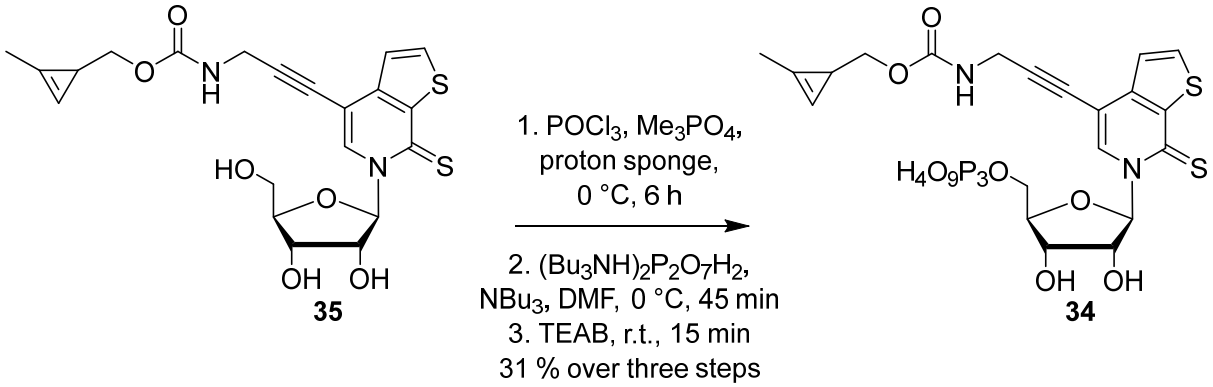
^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 3.00–2.96 (m; 12H; H-1), 1.70-1.62 (m; 12H; H-2), 1.36 (pseudo h; $^3J_{H\text{-}3/H\text{-}4} = {}^3J_{H\text{-}3/H\text{-}2}$ = 7.3 Hz; 12H; H-3), 0.93 (t; $^3J_{H\text{-}4/H\text{-}3}$ = 7.3 Hz; 18H; H-4).

^{31}P-NMR ($CDCl_3$, 162 MHz): δ [ppm] = -10.5 (s; P-α).



127

5.2.7.13 Synthesis of 2-(methylcycloprop-2-en-1-yl)methyl (((β-D-ribofuranos-5′-triphosphate-1′-yl)-7-thioxo-6,7-dihydrothieno[2,3-*c*]pyridin-4-yl)prop-2-yn-1-yl)carbamate (34, rTPT3^{CP} TP)



35 **34**

Experimental procedure:

Triphosphate synthesis was performed according to a protocol published by S. G. Srivatsan and Y. Tor.[130]

Nucleoside **35** (63 mg, 0.14 mmol, 1.0 eq.) and proton sponge (10 mg, 0.05 mmol, 0.3 eq.) were dried together under high vacuum overnight, dissolved in trimethyl phosphate (0.7 mL, freshly distilled and stored under argon and over 3 Å molecular sieve) and cooled in an ice bath. $POCl_3$ (40 µL, 0.42 mmol, 3.0 eq., freshly distilled and stored under argon) was added dropwise at 0 °C and the reaction was stirred under ice-cold conditions for 6 h. Then, a solution of bis-tri-*n*-butylammonium pyrophosphate (**127**) in dry DMF (0.5 M, 1.5 mL, 0.77 mmol, 5.5 eq.) and NBu_3 (0.35 mL, 1.47 mmol, 10.5 eq., stored over 3 Å molecular sieve overnight) were added simultaneously under ice-cooling. The mixture was stirred at 0 °C for 45 min and quenched by the addition of TEAB buffer (1.0 M, 10 mL, pH 7.5). The crude reaction was freeze-dried, redissolved in H_2O (5 mL) and purified via preparative HPLC (0→40 % MeCN/0.1 M TEAB in 6 min, flow = 40 mL min^{-1}). **34** was obtained as yellow solid (29.9 mg as 5-fold $HNEt_3^+$ salt, 24.7 µmol, 17 %).

Analytics:

Chemical Formula: $C_{21}H_{25}N_2O_{15}P_3S_2$.

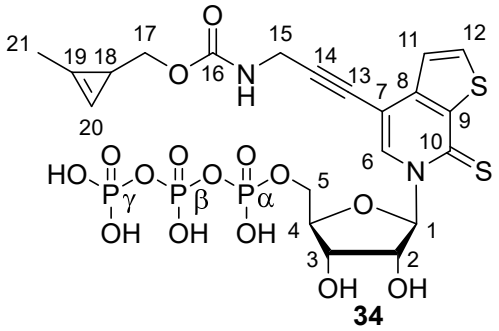
Molecular Weight [g mol^{-1}]: 702.48.

^{1}H-NMR (D_2O, 400 MHz): δ [ppm] = 8.37 (s; 1H; H-6), 8.05 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.4 Hz; 1H; H-12), 7.42 (m; 1H; H-11), 6.99 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 1.8 Hz; 1H; H-1), 6.64 (m; 1H; H-20), 4.52–4.34 (m; 5H; H-2, H-3, H-4, H-15), 4.23 (s; 2H; H-17), 4.10–3.94 (m; 2H; H-5), 2.09 (d, $^3J_{H\text{-}21/H\text{-}20}$ = 1.2 Hz; 3H; H-21), 1.65 (m; 1H; H-18).

^{31}P-NMR (D_2O, 162 MHz): δ [ppm] = -10.19 (d; $^2J_{P\text{-}\gamma/P\text{-}\beta}$ = 19.9 Hz; P-γ), -10.77 (d; $^2J_{P\text{-}\alpha/P\text{-}\beta}$ = 20.3 Hz; P-α), -22.58 (t; $^2J_{P\text{-}\beta/P\text{-}\alpha}$; $^2J_{P\text{-}\beta/P\text{-}\gamma}$ = 20.1 Hz; P-β).

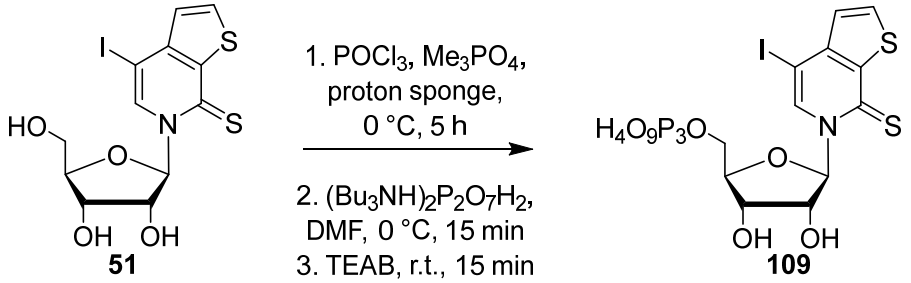
HR MS (ESI$^-$): calculated for $[M\text{-}H]^-$ ($C_{21}H_{25}N_2O_{15}P_3S_2^-$): 700.9836; found: *m*/*z* = 700.9843.

***t*$_R$ [min]:** 5.3; gradient: 5→100 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, flow 0.4 mL min^{-1}, *Zorbax* column.



34

5.2.7.14 Synthesis of (β-D-ribofuranos-5′-triphosphate-1′-yl)-4-iodo-6*H*-thieno[2,3-*c*]pyridine-7-thione (109, rTPT3^I TP)



51 **109**

Experimental procedure:

Triphosphate synthesis was performed according to a protocol published by S. G. Srivatsan and Y. Tor.[130]

Nucleoside **51** (41 mg, 0.10 mmol, 1.0 eq.) and proton sponge (10 mg, 0.05 mmol, 0.5 eq.) were dried together under high vacuum overnight, dissolved in trimethyl phosphate (0.6 mL, freshly distilled and stored under argon and over 3 Å molecular sieve) and cooled in an ice bath. $POCl_3$ (23 µL, 0.25 mmol, 2.5 eq., freshly distilled and stored under argon) was added

dropwise at 0 °C and the reaction was stirred under ice-cold conditions for 5 h. Then, a solution of bis-tributylammonium pyrophosphate (**127**) in dry DMF (0.5 M, 1.1 mL, 0.55 mmol, 5.5 eq.) and NBu_3 (0.25 mL, 1.05 mmol, 10.5 eq., stored over 3 Å molecular sieve overnight) were added simultaneously under ice-cooling. The mixture was stirred at 0 °C for 15 min and quenched by the addition of TEAB buffer (1.0 M, 7 mL, pH = 7.5). The crude reaction was freeze-dried, redissolved in H_2O (1.5 mL) and purified twice via preparative HPLC (0→40 % MeCN/0.1 M TEAA in 6 min, flow = 40 mL min^{-1}). **109** was obtained as yellow solid (5.4 mg as $HNEt_3^+$ salt, 7.1 µmol, 7 %).

Analytics:

Chemical Formula: $C_{12}H_{15}INO_{13}P_3S_2$.

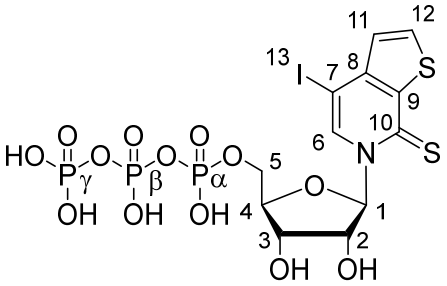
Molecular Weight [g mol^{-1}]: 665.20.

^{1}H-NMR (D_2O, 400 MHz): δ [ppm] = 8.47 (s; 1H; H-6), 8.05 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.4 Hz; 1H; H-12), 7.38 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.4 Hz; 1H; H-11), 6.97 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 2.3 Hz; 1H; H-1), 4.48–4.18 (m; 5H; H-2, H-3, H-4, H-5).

^{31}P-NMR (D_2O, 162 MHz): δ [ppm] = -9.17--9.21 (m; P-γ), -10.91 (d; $^2J_{P\text{-}\alpha/P\text{-}\beta}$ = 21.1 Hz; P-α), -22.23 (pseudo t; $^2J_{P\text{-}\beta/P\text{-}\alpha}$ = $^2J_{P\text{-}\beta/P\text{-}\gamma}$ = 20.1 Hz; P-β).

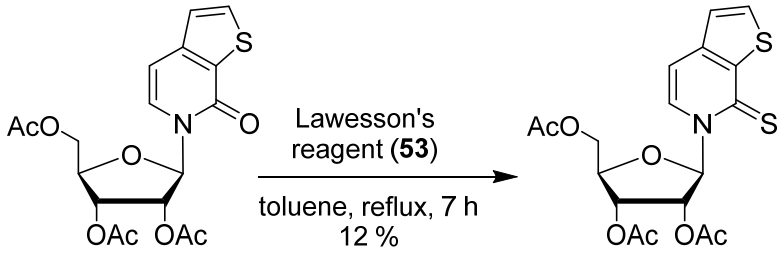
HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{12}H_{16}INO_{13}P_3S_2^+$): 665.8315; found: *m*/*z* = 665.8317.

***t*$_R$ [min]:** 5.1; gradient: 5→100 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, flow 0.4 mL min^{-1}, *Zorbax* column.



109

5.2.7.15 Synthesis of 6-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (116)



48 **116**

Experimental procedure:

Thienation of **48** was performed employing a description by Li *et al*.[194]

Under an atmosphere of argon **48** (2.00 g, 4.89 mmol, 1.0 eq.) and Lawesson's reagent (**53**, 1.98 g, 4.89 mmol, 1.0 eq.) were suspended in dry toluene (25 mL) and refluxed for 7 h. After cooling to room temperature the crude reaction mixture was evaporated to dryness, resuspended in dichloromethane (15 mL), filtered, and dried *in vacuo*. Purification via silica gel column chromatography (Cy/EtOAc, 1/1, R_f = 0.50) yielded product **116** (247 mg, 0.58 mmol, 12 %) as yellow solid.

Analytics (specific assignment given wherever possible):

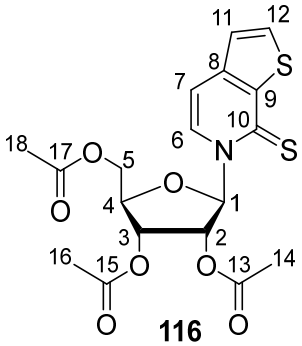
Chemical Formula: $C_{18}H_{19}NO_7S_2$.

Molecular Weight [g mol^{-1}]: 425.47.

^{1}H-NMR (CDCl$_3$, 500 MHz): δ [ppm] = 8.03 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.3 Hz; 1H; H-6), 7.83 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.3 Hz; 1H; H-12), 7.32 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 2.8 Hz; 1H; H-1), 7.28 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.3 Hz; 1H; H-11), 7.11 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.3 Hz; 1H; H-7), 5.57 (dd; $^3J_{H\text{-}2/H\text{-}3}$ = 5.3 Hz; $^3J_{H\text{-}2/H\text{-}1}$ = 2.8 Hz; 1H; H-2), 5.28 (dd; $^3J_{H\text{-}3/H\text{-}4}$ = 7.5 Hz; $^3J_{H\text{-}3/H\text{-}2}$ = 5.3 Hz; 1H; H-3), 4.53 (dt; $^3J_{H\text{-}4/H\text{-}3}$ = 7.5 Hz; $^3J_{H\text{-}4/H\text{-}5}$ = 3.0 Hz; 1H; H-4), 4.46 (d; $^3J_{H\text{-}5/H\text{-}4}$ = 3.0 Hz; 2H; H-5), 2.19 (s; 3H; H-14/16), 2.16 (s; 3H; H-16/14), 2.07 (s; 3H; H-18).

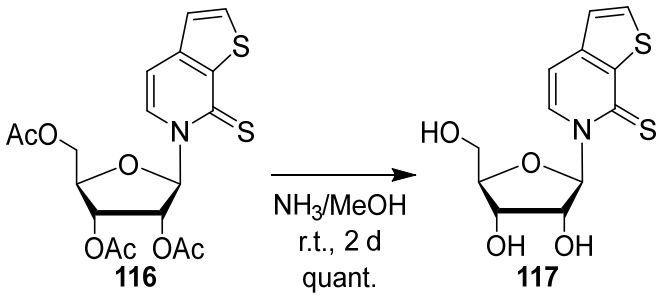
^{13}C-NMR (CDCl$_3$, 126 MHz): δ [ppm] = 174.5 (C-10), 170.2 (C-14/16/18), 169.6 (C-16/14/18), 169.3 (C18/16/14), 145.9 (C-8), 139.9 (C-9), 138.2 (C-12), 129.1 (C-6), 124.4 (C-11), 109.0 (C-7), 91.9 (C-1), 79.4 (C-4), 74.6 (C-2), 68.6 (C-3), 62.2 (C-5), 20.9 (C-14/16/18), 20.6 (C-16/14/18), 20.6 (C-18/16/14).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{18}H_{19}NO_7S_2Na^+$): 448.0495; found: *m*/*z* = 448.0499.



116

5.2.7.16 Synthesis of 6-(β-D-ribofuranos-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (117)



Experimental procedure:

Acyl deprotection took place following a protocol by Antonini *et al*.[314]

Under an atmosphere of argon a solution of NH_3 in MeOH (*c* = 7.0 M, 11.8 mL, 82.3 mmol, 500 eq.) was added to **116** (70 mg, 0.16 mmol, 1.0 eq.) at room temperature. The resulting solution was stirred for 43 h, dried under reduced pressure, and coevaporated with toluene (5 mL) several times. Nucleoside **117** (R_f = 0.39 in 95/5 CH_2Cl_2/MeOH, 49 mg, 0.16 mmol, quant.) was obtained as yellow solid.

Analytics:

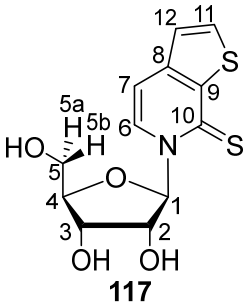
Chemical Formula: $C_{12}H_{13}NO_4S_2$.

Molecular Weight [g mol^{-1}]: 299.36.

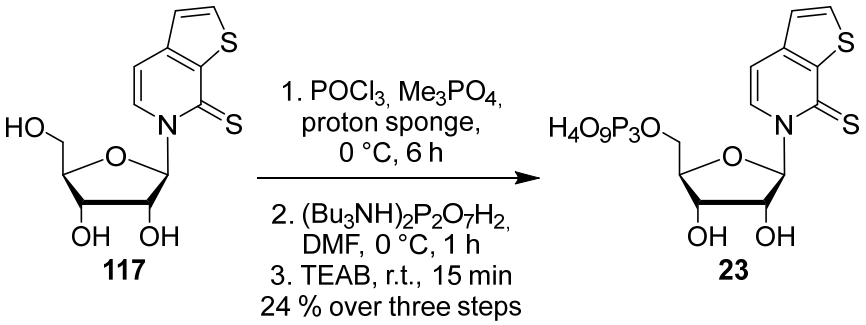
^{1}H-NMR (CD_3OD, 400 MHz): δ [ppm] = 8.59 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.2 Hz; 1H; H-6), 7.99 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.3 Hz; 1H; H-12), 7.38 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.3 Hz; 1H; H-11), 7.26 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.3 Hz; 1H; H-7), 7.04 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 1.8 Hz; 1H; H-1), 4.25 (dd; $^3J_{H\text{-}2/H\text{-}3}$ = 4.6 Hz; $^3J_{H\text{-}2/H\text{-}1}$ = 1.8 Hz; 1H; H-2), 4.23–4.15 (m; 2H; H-4; H-3), 4.06 (dd, $^3J_{H\text{-}5a/H\text{-}5b}$ = 12.5 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 2.2 Hz; 1H; H-5a/5b), 3.88 (dd; $^3J_{H\text{-}5b/H\text{-}5a}$ = 12.6 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 2.7 Hz; 1H; H-5b/5a).

^{13}C-NMR (CD_3OD, 101 MHz): δ [ppm] = 174.3 (C-10), 146.3 (C-9), 141.0 (C-8), 139.0 (C-12), 132.3 (C-6), 125.6 (C-11), 109.8 (C-7), 96.3 (C-1), 85.7 (C-4), 77.5 (C-2), 69.5 (C-3), 61.0 (C-5).

HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{12}H_{14}NO_4S_2^+$): 300.0359; found: *m*/*z* = 300.0355.



5.2.7.17 Synthesis of 6-(β-D-ribofuranos-5′-triphosphate-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (rTPT3 TP, 23)



Experimental procedure:

Generation of ribosyl 5′ triphosphate derivative **23** was performed according to a protocol published by S. G. Srivatsan and Y. Tor.[130]

Nucleoside **117** (19 mg, 63 µmol, 1.0 eq.) and proton sponge (14 mg, 65 µmol, 1.0 eq.) were dried under high vacuum for 16 h. Under an atmosphere of argon, freshly distilled trimethyl phosphate (stored over molecular sieve 3 Å, 300 µL) was added and the solution was cooled in an ice bath. $POCl_3$ (freshly distilled and stored under argon, 18 µL, 190 µmol, 3.0 eq.) was added dropwise at 0 °C. The mixture was stirred for 6 h at 0 °C, subsequently tributylamine (160 µL, 670 µmol, 11.0 eq.) and a solution of bis-tributylammonium pyrophosphate in dry DMF (0.5 M, 700 µL, 350 µmol, 5.5 eq.) were added simultaneously. After 60 min of stirring under ice-cold conditions, the reaction was quenched by the addition of TEAB (1.0 M, 7.0 mL). The mixture was dried *in vacuo* using a rotary evaporator connected to a high vacuum pump and co-evaporated with water (3×10 mL). After redissolving in ddH_2O (1.5 mL) the crude product was purified via preparative HPLC (0→40 % MeCN/0.1 % TEAB pH 7.5 in 6 min, flow rate 40 mL min^{-1}). Triphosphate **rTPT3** TP (**23**) was obtained as a faint yellow solid (8.0 mg as 2-fold NEt_3 salt, 14.8 µmol, 24 %).

Analytics:

Chemical Formula: $C_{12}H_{16}NO_{13}P_3S_2$.

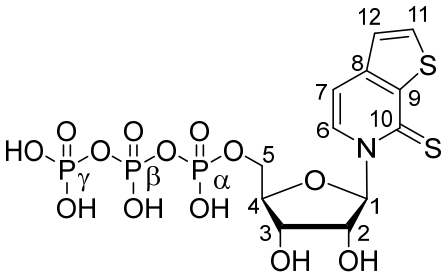
Molecular Weight [g mol^{-1}]: 539.30.

^{1}H-NMR (D_2O, 400 MHz): δ [ppm] = 8.52 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.2 Hz; 1H; H-6), 8.11 (d; $^3J_{H\text{-}12/H\text{-}11}$ = 5.4 Hz; 1H; H-12), 7.62 (d, $^3J_{H\text{-}7/H\text{-}6}$ = 7.2 Hz; 1H; H-7), 7.54 (d; $^3J_{H\text{-}11/H\text{-}12}$ = 5.4 Hz; 1H; H-11), 7.18 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 2.2 Hz; 1H; H-1), 4.54–4.38 (m; 5H; H-2, H-3, H-4, H-5).

^{31}P-NMR (D_2O, 162 MHz): δ [ppm] = -9.54 (m; P-γ), -11.44 (d; $^3J_{P\text{-}\alpha/P\text{-}\beta}$ = 20.0 Hz; P-α), -23.04 (pseudo t; $^3J_{P\text{-}\beta/P\text{-}\alpha}$ = $^3J_{P\text{-}\beta/P\text{-}\gamma}$ = 20.0 Hz; P-β).

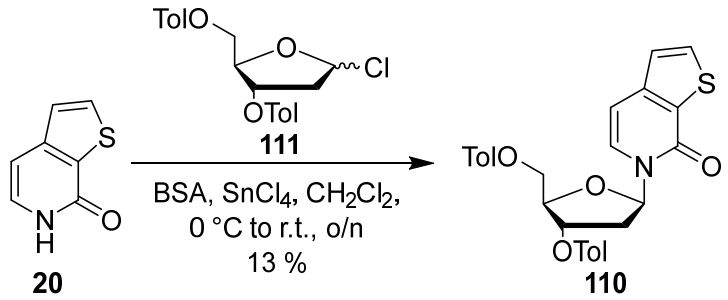
HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{12}H_{17}NO_{13}P_3S_2^+$): 539.9349; found: *m*/*z* = 539.9354.

***t*$_R$ [min]:** 2.0; gradient: 5→100 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, flow 0.4 mL min^{-1}, *Zorbax* column.



23

5.2.7.18 Synthesis of 6-(3′,5′-toluoyldeoxyribofuranos-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-one (110)



Experimental procedure:

Synthesis of toluoyl protected deoxy nucleoside analog **110** was conducted as described by Li *et al*.[194]

Under an atmosphere of argon pyridone **20** (1.62 g, 10.7 mmol, 1.0 eq.) was dissolved in dry CH_2Cl_2 (70 mL). *N*,*O*-Bis(trimethylsilyl)acetamide (BSA, 3.2 mL, 12.9 mmol, 1.2 eq.) was added and the mixture was stirred for 1 h at room temperature. Then toluoyl protected deoxyribose chloride **111** (5.00 mg, 12.9 mmol, 1.2 eq.) was added and the solution was cooled in an ice bath. After the dropwise addition of $SnCl_4$ (1.3 mL, 10.7 mmol, 1.0 eq.) the reaction was allowed to warm to room temperature and stirred overnight. Quenching with saturated aqueous $NaHCO_3$ solution (350 mL) was followed by extraction with EtOAc (3×200 mL), drying over Na_2SO_4, and evaporation. The crude product was purified via column chromatography (Cy/EtOAc, 6/4, R_f = 0.54) yielding **110** (712 mg, 1.41 mmol, 13 %) as an off-white foam.

Analytics (specific assignment given wherever possible):

Chemical Formula: $C_{28}H_{25}NO_6S$.

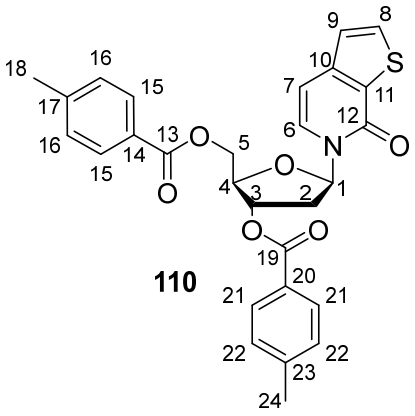
Molecular Weight [g mol^{-1}]: 503.57.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 7.97 (d; $^3J_{H\text{-}21/H\text{-}22}$ = 8.2 Hz; 2H; H-21/15), 7.91 (d, $^3J_{H\text{-}15/H\text{-}16}$ = 8.2 Hz; 2H; H-15/21), 7.69 (d; $^3J_{H\text{-}8/H\text{-}9}$ = 5.1 Hz; 1H; H-8), 7.55 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.4 Hz; 1H; H-6), 7.27 (d; $^3J_{H\text{-}22/H\text{-}21}$ = 8.1 Hz; 2H; H-22/16), 7.21 (d; $^3J_{H\text{-}16/H\text{-}15}$ = 8.2 Hz; 2H; H-16/22), 7.16 (d; $^3J_{H\text{-}9/H\text{-}8}$ = 5.2 Hz; 1H; H-9), 6.85 (dd;

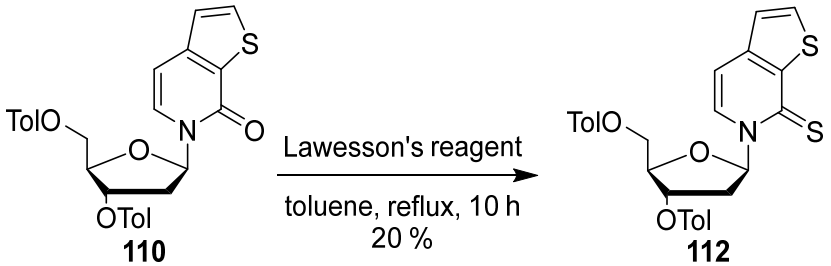
$^{3}J_{H\text{-}1/H\text{-}2a}$ = 8.4 Hz; $^{3}J_{H\text{-}1/H\text{-}2b}$ = 5.6 Hz; 1H; H-1), 6.57 (d; $^{3}J_{H\text{-}7/H\text{-}6}$ = 7.4 Hz; 1H; H-7), 5.65 (m; 1H; H-4), 4.76 (dd; $^{2}J_{H\text{-}5a/H\text{-}5b}$ = 12.1 Hz; $^{3}J_{H\text{-}5a/H\text{-}4}$ = 3.4 Hz; 1H; H-5a/5b), 4.71 (dd; $^{2}J_{H\text{-}5b/H\text{-}5a}$ = 12.1 Hz; $^{3}J_{H\text{-}5b/H\text{-}4}$ = 3.6 Hz; 1H; H-5b/5a), 4.61 (m; 1H; H-3), 2.92 (ddd; $^{2}J_{H\text{-}2a/H\text{-}2b}$ = 14.4 Hz; $^{3}J_{H\text{-}2a/H\text{-}1}$ = 5.6 Hz; $^{3}J_{H\text{-}2a/H\text{-}3}$ = 1.9 Hz; 1H; H-2a), 2.43 (s; 3H; H-24/18), 2.40 (s; 3H; H-18/24) 2.36 (ddd; $^{2}J_{H\text{-}2b/H\text{-}2a}$ = 14.6 Hz; $^{3}J_{H\text{-}2b/H\text{-}1}$ = 8.6 Hz; $^{3}J_{H\text{-}2b/H\text{-}3}$ = 6.6 Hz; 1H; H-2b/2a).

^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 166.3 (C-19/13), 166.2 (C-13/19), 158.2 (C-12), 145.2 (C-8), 144.5 (C-21/15), 144.3 (C-15/21), 133.9 (C-8), 130.1 (C-9), 130.0 (C-21/15), 129.7 (C-15/21), 129.4 (C-22/16), 129.4 (C-16/22), 127.0 (C-6), 126.8 (C-22/16), 126.6 (C-16/22), 124.3 (C-9), 103.7 (C-7), 85.6 (C-1), 83.0 (C-3), 75.2 (C-4), 64.5 (C-5), 39.3 (C-2), 21.9 (C-24/18), 21.8 (C-18/24).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{28}H_{25}NO_6SNa^+$): 526.1295; found: *m*/*z* = 526.1292.



5.2.7.19 Synthesis of 6-(3′,5′-toluoyldeoxyribofuranos-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (112)



Experimental procedure:

Thienation was achieved adapting a protocol by Li *et al*.[194]

Compound **110** was dried by three coevaporations with dry toluene (2 mL). Under an atmosphere of argon a solution of **110** (70 mg, 0.14 mmol, 1.0 eq.) in dry toluene (1.5 mL) was added to Lawesson's reagent (**53**, 525 mg, 1.30 mmol, 1.1 eq.) and iteratively refluxed for a total of 15 h (3×5 h). After cooling down to ambient temperatures, the crude reaction

mixture was filtrated and evaporated before it was chromatographed (Cy/EtOAc, 6/4, R_f = 0.66) to give **112** (20 mg, 0.04 mmol, 29 %) as yellow solid.

Analytics (specific assignment given wherever possible):

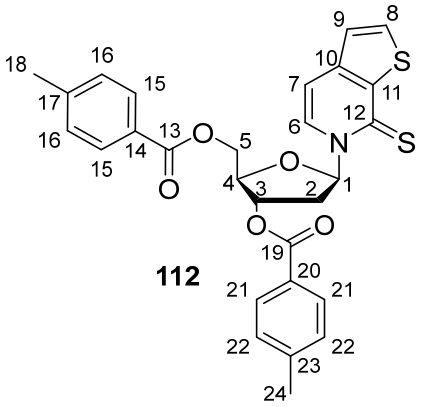
Chemical Formula: $C_{28}H_{25}NO_5S_2$.

Molecular Weight [g mol^{-1}]: 519.63.

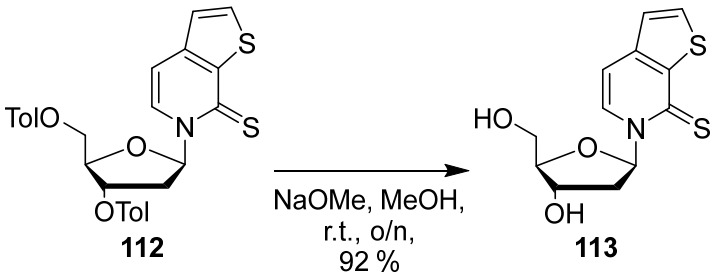
^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 8.09 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.2 Hz; 1H; H-6), 8.01–7.96 (m; 2H; H-21/15), 7.92-7.86 (m; 2H; H-15/21), 7.80 (d; $^3J_{H\text{-}8/H\text{-}9}$ = 5.3 Hz; 1H; H-8), 7.48 (dd; $^3J_{H\text{-}1/H\text{-}2a}$ = 7.6 Hz; $^3J_{H\text{-}1/H\text{-}2b}$ = 5.8 Hz; 1H; H-1), 7.31–7.27 (m; 2H; H-22/16), 7.24 (d; $^3J_{H\text{-}9/H\text{-}8}$ = 5.3 Hz; 1H; H-9), 7.21 (d; $^3J_{H\text{-}16/H\text{-}15}$ = 8.1 Hz; 2H; H-16/22), 6.97 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.3 Hz; 1H; H-7), 5.65 (m; 1H; H-4), 4.83 (dd; $^2J_{H\text{-}5b/H\text{-}5a}$ = 12.3 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 3.1 Hz; 1H; H-5b/5a), 4.77 (dd; $^2J_{H\text{-}5a/H\text{-}5b}$ = 12.3 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 3.5 Hz; 1H; H-5a/5b), 4.69 (m; 1H; H-3), 3.33 (ddd; $^2J_{H\text{-}2a/H2b}$ = 14.5 Hz; $^3J_{H\text{-}2a/H\text{-}1}$ = 5.8 Hz; $^3J_{H\text{-}2a/H\text{-}3}$ = 2.4 Hz; 1H; H-2a/2b), 2.44 (s; 3H; H-24/18), 2.40 (s; 3H; H-18/24), 2.30 (m; 1H; H-2b/2a).

^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 173.2 (C-12), 166.3 (C-14, C-20), 145.7 (C-10), 144.6 (C-19/13), 144.5 (C-13/19), 139.2 (C-11), 137.8 (C-8), 130.0 (C-21/15), 129.7 (C-15/21), 129.5 (C-22/16), 129.4 (C-16/22), 129.2 (C-6), 126.7 (C-23/17), 126.5 (C-17/23), 124.4 (C-9), 109.2 (C-7), 91.1 (C-1), 83.8 (C-3), 74.7 (C-4), 64.3 (C-5), 39.1 (C-2), 21.9 (C-.24/18), 21.8 (C-18/24).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{28}H_{25}NO_5S_2Na^+$): 542.1066; found: *m*/*z* = 542.1069.



5.2.7.20 Synthesis of 6-(3′,5′-deoxyribofuranos-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (113)



Experimental procedure:

Toluoyl deprotection took place according to Li *et al*.[194]

Toluoyl protected deoxynucleoside **112** (20 mg, $3.85 \cdot 10^{-2}$ mmol, 1.0 eq.) was dissolved in dry methanol (1 mL) and a large excess of a methanolic NaOMe solution (5.4 M, 1 mL, 5.1 mmol 140 eq.) was added dropwise at room temperature. The mixture was stirred overnight, concentrated *in vacuo* and chromatographed (CH_2Cl_2/MeOH, 95/5, $R_f = 0.65$) to yield nucleoside **113** (10 mg, $3.53 \cdot 10^{-2}$ mmol, 92 %) as dark yellow solid.

Analytics:

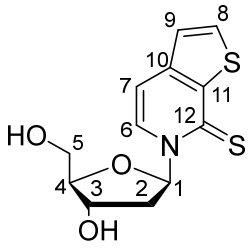
Chemical Formula: $C_{12}H_{13}NO_3S_2$.

Molecular Weight [g mol^{-1}]: 283.37.

^{1}H-NMR (CD_3OD, 400 MHz): δ [ppm] = 8.46 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.2 Hz; 1H; H-6), 7.99 (d; $^3J_{H\text{-}8/H\text{-}9}$ = 5.3 Hz; 1H; H-8), 7.38 (d; $^3J_{H\text{-}9/H\text{-}8}$ = 5.4 Hz; 1H; H-9), 7.35 (pseudo t; $^3J_{H\text{-}1/H\text{-}2a} = {}^3J_{H\text{-}1/H\text{-}2b}$ = 6.3 Hz; 1H; H-1), 7.28 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.2 Hz; 1H; H-7), 4.44 (pseudo dt; $^3J_{H\text{-}3/H\text{-}2a}$ = 6.4 Hz; $^3J_{H\text{-}3/H\text{-}2b} = {}^3J_{H\text{-}3/H\text{-}4}$ = 4.2 Hz; 1H; H-3), 4.07 (pseudo q; $^3J_{H\text{-}4/H\text{-}5b} = {}^3J_{H\text{-}4/H\text{-}5a} = {}^3J_{H\text{-}4/H\text{-}3}$ = 3.8 Hz; 1H; H-4), 3.93 (dd; $^3J_{H\text{-}5a/H\text{-}5b}$ = 12.1 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 3.2 Hz; 1H; H-5a), 3.84 (dd; $^3J_{H\text{-}5b/H\text{-}5a}$ = 12.1 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 3.8 Hz; 1H; H-5b), 2.75 (ddd; $^3J_{H\text{-}2b/H\text{-}2a}$ = 13.7 Hz; $^3J_{H\text{-}2b/H\text{-}1}$ = 6.3 Hz; $^3J_{H\text{-}2b/H\text{-}3}$ = 4.4 Hz; 1H, H-2b), 2.14 (pseudo dt; $^3J_{H\text{-}2a/H\text{-}2b}$ = 13.7 Hz; $^3J_{H\text{-}2a/H\text{-}3} = {}^3J_{H\text{-}2a/H\text{-}1}$ = 6.3 Hz; 1H; H-2a).

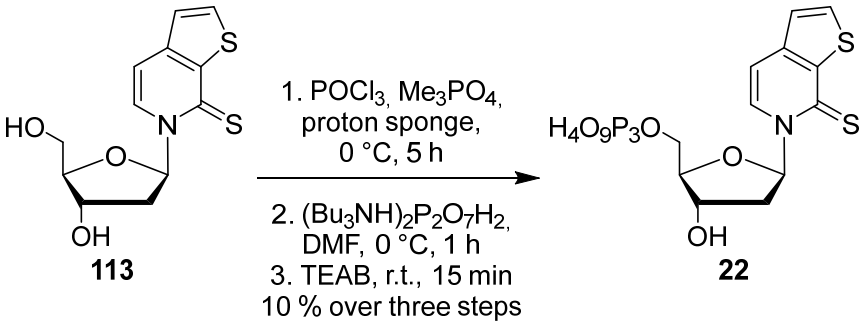
^{13}C-NMR (CD_3OD, 101 MHz): *δ* [ppm] = 173.7 (C-12), 146.2 (C-10), 141.1 (C-11), 139.0 (C-8), 131.9 (C-6), 125.6 (C-9), 110.1 (C-7), 92.1 (C-1), 89.6 (C-4), 71.5 (C-3), 62.4 (C-5), 42.7 (C-2).

HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{12}H_{14}NO_3S_2^+$): 284.0410; found: *m*/*z* = 284.0409.



113

5.2.7.21 Synthesis of 6-(β-D-Deoxyribofuranos-5′-triphosphate-1-yl)-6*H*-thieno[2,3-*c*]pyridin-7-thione (dTPT3 TP, 22)



Experimental procedure:

Triphosphate synthesis was performed according to a protocol published by S. G. Srivatsan and Y. Tor.[130]

Nucleoside **113** (40 mg, 140 µmol, 1.0 eq.) and proton sponge (30 mg, 140 µmol, 1.0 eq.) were dried under high vacuum overnight. Under an atmosphere of argon freshly distilled trimethyl phosphate (stored over molecular sieve 3 Å and under argon, 500 µL) was added and the solution was cooled in an ice bath. $POCl_3$ (freshly distilled and stored under argon, 39 µL, 420 µmol, 3.0 eq.) was added dropwise at 0 °C. The mixture was stirred for 5 h at 0 °C, subsequently tributylamine (350 µL, 1.48 mmol, 10.5 eq.) and a solution of bis-tributylammonium pyrophosphate in dry DMF (0.5 M, 1.56 mL, 0.78 mmol, 5.5 eq.) were added simultaneously. After 1 h of stirring under ice-cold conditions, the reaction was quenched by the addition of TEAB (1.0 M, 10 mL). The mixture was dried *in vacuo* using a rotary evaporator connected to a high vacuum pump, co-evaporated with water (3×10 mL), and freeze-dried overnight. Dissolved in ddH_2O (3.5 mL) the crude product was purified via preparative HPLC (0→40 % MeCN/0.1 % TEAB pH 7.5 in 6 min, flow rate 40 mL min^{-1}) and analytical HPLC (10→40 % MeCN/0.1 % TEAB pH 7.5 in 20 min, flow rate 1 mL min^{-1}). Triphosphate **22** (2.2 mg as 2-fold NEt_3 salt, 3.03 µmol, 2 % and 10.9 mg als 4-fold NEt_3 salt, 11.7 µmol, 8 %) was obtained as a faint yellow solid.

Analytics:

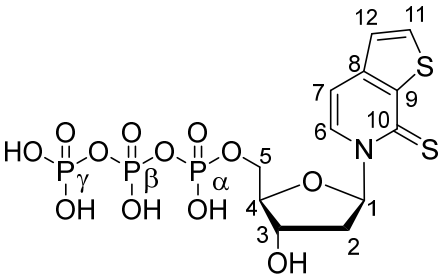
Chemical Formula: $C_{12}H_{16}NO_{12}S_2$.

Molecular Weight [g mol^{-1}]: 523.30.

^{31}P NMR (D_2O, 162 MHz): δ [ppm] = -10.35 (m, P-γ), -11.38 (m, P-α), (pseudo t, $^3J_{P\text{-}\beta/P\text{-}\alpha} = {}^3J_{P\text{-}\beta/P\text{-}\gamma}$ = 19.9 Hz, P-β).

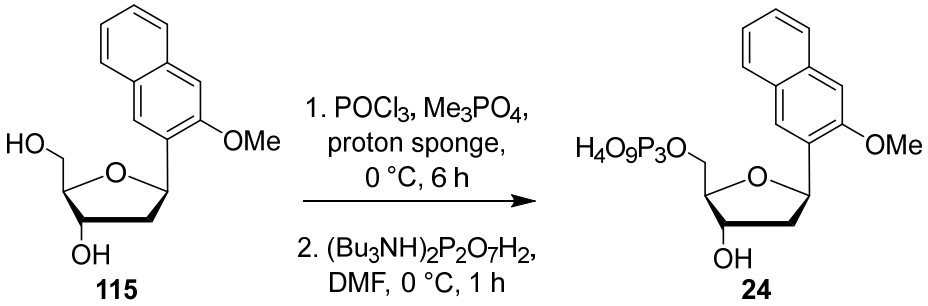
HR MS (ESI$^-$): calculated for $[M\text{-}H]^-$ ($C_{12}H_{15}NO_{12}P_3S_2^-$): 521.9254; found: *m*/*z* = 521.9197.

***t*$_R$ [min]:** 1.0; gradient: 5→100 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, flow 0.4 mL min^{-1}, *Zorbax* column.



22

5.2.7.22 Synthesis of 1-(β-D-Deoxyribofuranos-5′-triphosphate-1-yl)-3-methoxy-2-naphthalene (dNaM TP, 24)



115 **24**

Experimental procedure:

Triphosphate synthesis was adapted from S. G. Srivatsan and Y. Tor.[130]

Nucleoside **115** (25 mg, 91 µmol, 1.0 eq.) and proton sponge (20 mg, 91 µmol, 1.0 eq.) were dried under high vacuum overnight. Under an atmosphere of argon freshly distilled Me_3PO_4 (stored over molecular sieve 3 Å and under argon, 500 µL) was added and the solution was cooled in an ice bath. $POCl_3$ (freshly distilled and stored under argon, 25 µL, 270 µmol, 3.0 eq.) was added dropwise at 0 °C. The mixture was stirred for 6 h at 0 °C, subsequently tributylamine (230 µL, 0.96 mmol, 10.5 eq.) and a solution of bis-tributylammonium pyrophosphate in dry DMF (0.5 M, 1.00 mL, 0.50 mmol, 5.5 eq.) were added simultaneously. After 1 h of stirring under ice-cold conditions, the reaction was quenched by the addition of TEAB (1.0 M, 7 mL). The crude reaction mixture was evaporated and co-evaporated with water (3×10 mL), and freeze-dried overnight. Dissolved in ddH_2O (3.0 mL) the crude product was purified via preparative HPLC (0→40 % MeCN/0.1 % TEAB pH 7.5 in 6 min, flow rate 40 mL min^{-1}) to give triphosphate **24** (2.9 mg as 2-fold NEt_3 salt, 4 µmol, 4 %) was obtained as white solid.

Analytics:

Chemical Formula: $C_{16}H_{21}O_{13}P_3$.

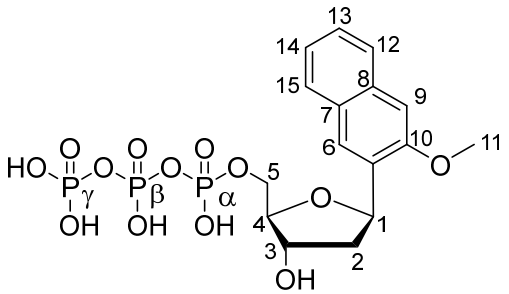
Molecular Weight [g mol^{-1}]: 514.25.

^{1}H-NMR (D_2O, 400 MHz): δ [ppm] = 8.08 (s; 1H; H-9), 7.97 (d; $^3J_{H\text{-}12/H\text{-}13}$ = 8.1 Hz; 1H; H-12), 7.88 (d; $^3J_{H\text{-}15/H\text{-}14}$ = 8.6 Hz; 1H; H-15), 7.54 (ddd; $^3J_{H\text{-}14/H\text{-}15}$ = 8.3 Hz; $^3J_{H\text{-}14/H\text{-}13}$ = 6.9 Hz; $^4J_{H\text{-}14/H\text{-}12}$ = 1.3 Hz; 1H; H-14), 7.46 (ddd; $^3J_{H\text{-}13/H\text{-}12}$ = 8.1 Hz; $^3J_{H\text{-}13/H\text{-}14}$ = 6.9 Hz; $^3J_{H\text{-}13/H\text{-}15}$ = 1.3 Hz; 1H; H-13), 7.41 (s; 1H; H-6), 5.61 (dd; $^3J_{H\text{-}1/H\text{-}2a}$ = 10.0 Hz; $^3J_{H\text{-}1/H\text{-}2b}$ = 5.8 Hz; 1H; H-1), 4.64 (m; 1H; H-5a/5b), 4.28 (m; 1H; H-5b/5a), 4.22–4.17 (m; 2H; H-3; H-4), 4.00 (s; 3H; H-11) 2.45 (ddd; $^2J_{H\text{-}2b/H\text{-}2a}$ = 13.5 Hz; $^3J_{H\text{-}2b/H\text{-}1}$ = 5.9 Hz; $^3J_{H\text{-}2b/H\text{-}3}$ = 2.3 Hz; 1H; H-2b/2a), 2.24 (ddd; $^2J_{H\text{-}2a/H\text{-}2b}$ = 13.5 Hz; $^3J_{H\text{-}2a/H\text{-}1}$ = 10.1 Hz; $^3J_{H\text{-}2a/H\text{-}3}$ = 6.0 Hz; 1H; H-2a/2b).

^{31}P NMR (D_2O, 162 MHz): δ [ppm] = -7.05 (m; P-γ), -11.86 (d; $^3J_{P\text{-}\alpha/P\text{-}\beta}$ = 19.4 Hz; P-α), -22.79 (pseudo t; $^3J_{P\text{-}\beta/P\text{-}\alpha}$ = $^3J_{P\text{-}\beta/P\text{-}\gamma}$ = 19.9 Hz; P-β).

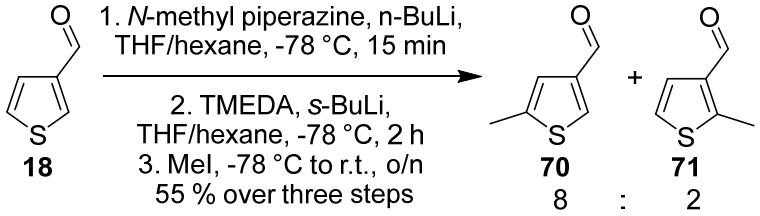
HR MS (ESI^-): calculated for $[M\text{-}H]^-$ ($C_{16}H_{20}O_{13}P_3^-$): 513.0122; found: *m*/*z* = 513.0100.

***t*$_R$ [min]:** 6.0; gradient: 0→60 % MeCN/0.1 % (*w*/*v*) NH_4OAc in 20 min, flow 0.4 mL min^{-1}, *Zorbax* column.



24

5.2.7.23 Synthesis of 5-methylthiophene-3-carboxaldehyde (70)



Experimental procedure:

Methylation of thiophene **18** was achieved following a protocol by D. L. Comins and M. O. Killpack.[397]

Under an atmosphere of argon *N*-methylpiperazine (8.3 mL, 75 mmol, 1.2 eq.) was dissolved in dry THF (20 mL) and cooled to -78 °C. *n*-BuLi (2.5 M in hexane, 30 mL, 75 mmol, 1.2 eq.) was added dropwise, the reaction was stirred for 15 min and then thiophene-3-carboxaldehyde (**18**, 5.7 mL, 62 mmol, 1.0 eq.) was added. After stirring at 78 °C for 15 min TMEDA (18.6 mL, 125 mmol, 2.0 eq.) and *sec*-BuLi (1.3 M in hexane, 58 mL, 75 mmol, 1.2 eq.) were added consecutively. The reaction mixture was stirred at -78 °C for 2 h before

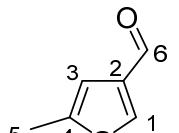
the dropwise addition of MeI (15.5 mL, 250 mmol, 4.0. eq.) The reaction mixture was allowed to reach room temperature, stirred overnight, and poured into ice-cold water (250 mL). The aqueous layer was extracted with Et_2O (3×200 mL), dried over Na_2SO_4, evaporated, and purified via column chromatography (Cy/EtOAc 5/1, R_f = 0.63) to give product **70** (6.46 g as 65 %:17 %:18 % mixture with starting material and byproduct **71**, 34 mmol, 53 %) as a faint yellow liquid.

Analytics:

Chemical Formula: C_6H_6OS.

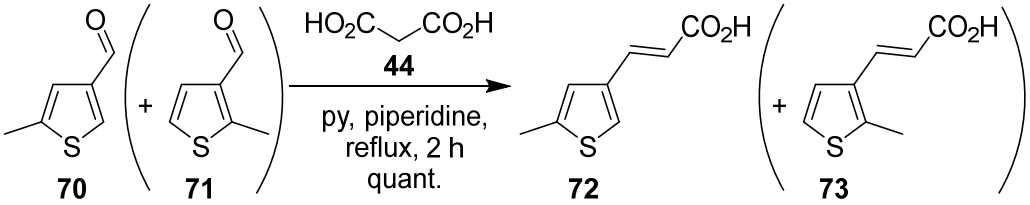
Molecular Weight [g mol^{-1}]: 126.18.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = (s; 1H; H-6), 7.87 (d; $^3J_{H\text{-}3/H\text{-}5}$ = 1.4 Hz; 1H; H-3), 7.17 (s; 1H; H-1), 2.49 (d; $^3J_{H\text{-}5/H\text{-}3}$ = 1.2 Hz; 3H; H-5).



70

5.2.7.24 Synthesis of (*2E*)-3-(5-methyl-thiophen-3-yl)-acrylic acid (72)



Experimental procedure:

Knoevenagel condensation was performed according to New *et al.*[306]

Aldehyde **70** (ca. 65:17:18 mixture with 2-methyl isomer **71** and **18**, 5.1 g, 40 mmol, 1.0 eq.), malonic acid (**44**, 4.2 g, 40 mmol, 1.0 eq.), and piperidine (2.0 mL, 20 mmol, 0.5 eq.) were dissolved in pyridine (50 mL) and refluxed for 2 h. After cooling to ambient temperatures the mixture was carefully poured into cold HCl conc. (100 mL) under ice cooling. The resulting suspension was extracted with EtOAc (3×250 mL), the organic layers were dried over Na_2SO_4 and evaporated. Acrylic acid **72** (6.68 g as ca. 5:1:1 mixture with 2-methyl isomer **73** and **19**, 40 mmol, quant.) was obtained as beige solid.

Analytics:

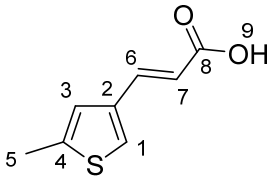
Chemical Formula: $C_8H_8O_2S$.

Molecular Weight [g mol^{-1}]: 168.21.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 11.36 (br s; 1H; H-9), 7.67 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 15.8 Hz; 1H; H-6), 7.31 (d; $^4J_{H\text{-}1/H\text{-}3}$ = 1.4 Hz; 1H; H-1), 6.98 (m; 1H; H-3), 6.18 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 15.8 Hz; 1H; H-7), 2.49 (d; $^4J_{H\text{-}5/H\text{-}3}$ = 1.2 Hz; 3H; H-5).

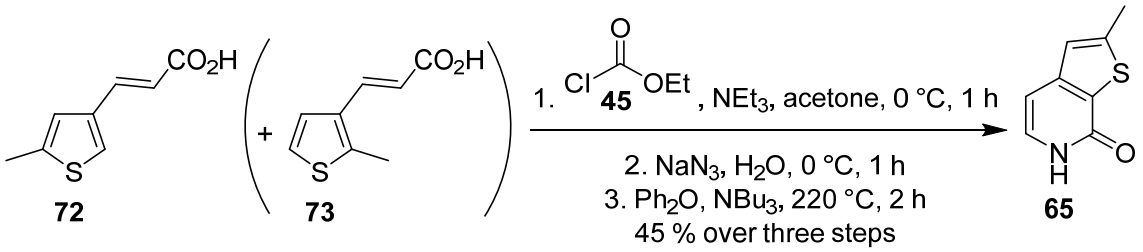
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 173.2 (C-8), 141.8 (C-4), 140.9 (C-6), 137.4 (C-2), 127.9 (C-1/3), 123.1 (C-3/1), 116.5 (C-6), 15.5 (C-5).

HR MS (ESI$^-$): calculated for $[M-H]^-$ ($C_8H_7O_2S^-$): 167.0172; found: *m*/*z* = 167.0173.



72

5.2.7.25 Synthesis of 2-methyl-6*H*-thieno[2,3-*c*]pyridin-7-one (65)



Experimental procedure:

Pyridone **65** was prepared adapting a procedure published by Ewing *et al*.[307]

Acrylic acid **72** (ca. 5:1:1 mixture with 2-methyl isomer **73** and **19**, 2.4 g, 14.3 mmol, 1.0 eq.) and NEt_3 (2.2 mL, 15.7 mmol, 1.1 eq.) were dissolved in acetone (70 mL) and cooled in an ice bath. Ethyl chloroformate (1.5 mL, 15.7 mmol, 1.1 eq.) was added dropwise and the reaction was stirred for 1 h at 0 °C. A solution of NaN_3 (1.4 g, 21.4 mmol, 1.5 eq.) in water (10 mL) was added and the mixture stirred for further 1 h under ice cooling. The reaction was poured in cold water (250 mL), the precipitated acid azide was collected by filtration and freeze-dried overnight. Suspended in warm Ph_2O (20 mL, ~40 °C) the slurry was added portionwise to NBu_3 (3.7 mL, 15.7 mmol, 1.1 eq.) in Ph_2O (50 mL) at 220 °C and stirred for 2 h before cooling to room temperature. Collection by filtration and drying under high vacuum yielded methyl pyridone derivative **65** (1.07 g, 6.5 mmol, 45 %) as a beige powder.

Analytics:

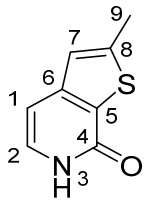
Chemical Formula: C_8H_7NOS.

Molecular Weight [g mol^{-1}]: 165.21.

^{1}H-NMR (DMSO-d_6, 400 MHz): δ [ppm] = 11.38 (br s; 1H; H-3), 7.21 (d; $^3J_{H\text{-}2/H\text{-}1}$ = 6.9 Hz; 1H; H-2), 7.05 (d; $^4J_{H\text{-}8/H\text{-}9}$ = 1.4 Hz; 1H; H-8), 6.57 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 6.9 Hz; 1H; H-1), 2.55 (d; $^4J_{H\text{-}9/H\text{-}8}$ = 1.1 Hz; 3H; H-9).

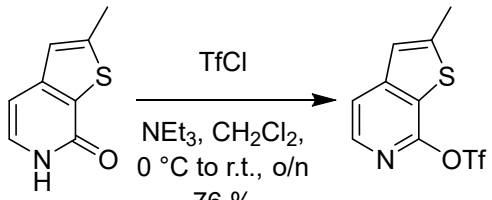
^{13}C-NMR (DMSO-d_6, 101 MHz): δ [ppm] = 158.0 (C-4), 147.7 (C-6/8), 146.8 (C-8/6), 130.2 (C-2), 127.6 (C-5), 123.2 (C-7), 101.7 (C-1), 15.8 (C-9).

HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_8H_8NOS^+$): 166.0321; found: *m*/*z* = 166.0320.



65

5.2.7.26 Synthesis of trifluoro-methanesulfonic acid 2-methylthieno[2,3-*c*]pyridine-7-yl ester (75)



65 **75**

Experimental procedure:

Under an atmosphere of argon **65** (200 mg, 1.21 mmol, 1.0 eq.) was suspended in dry CH_2Cl_2 (15 mL) and NEt_3 (0.84 mL, 6.05 mmol, 5.0 eq.) and cooled in an ice bath. TfCl (1.29 mL, 12.1 mmol, 10.0 eq.) was added dropwise, the reaction was allowed to warm to ambient temperature and stirred overnight. After the addition of water (15 mL) the aqueous phase was extracted with CH_2Cl_2 (3×10 mL). The combined organic layers were dried over Na_2SO_4 and evaporated. Column chromatographic purification gave **75** (268 mg, 0.92 mmol, 76 %) as a slowly crystallizing brown oil.

Analytics:

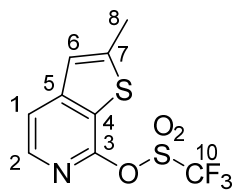
Chemical Formula: $C_9H_6F_3NO_3S_2$.

Molecular Weight [g mol^{-1}]: 297.27.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 7.53 (d; $^3J_{H\text{-}2/H\text{-}1}$ = 7.8 Hz; 1H; H-2), 6.91 (q; $^4J_{H\text{-}6/H\text{-}7}$ = 1.1 Hz; 1H; H-6), 6.64 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 7.7 Hz; 1H; H-1), 2.63 (d; $^4J_{H\text{-}8/H\text{-}6}$ = 1.1 Hz; 3H; H-8).

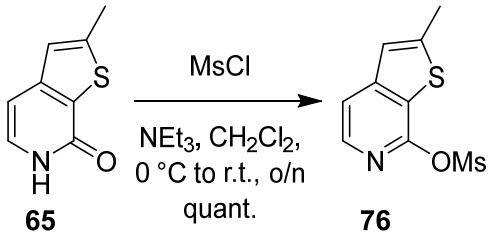
^{19}F-NMR ($CDCl_3$, 470 MHz): δ [ppm] = -70.6 (F-10).

HR MS (EI): calculated for $[M]^{\bullet+}$ ($C_9H_6F_3NO_3S_2^{\bullet+}$): 296.9736; found: *m*/*z* = 296.9742.



75

5.2.7.27 Synthesis of methanesulfonic acid 2-methylthieno[2,3-*c*]pyridine-7-yl ester (76)



65 **76**

Experimental procedure:

65 (100 mg, 0.61 mmol, 1.0 eq.) was suspended in dry CH_2Cl_2 (10 mL) and NEt_3 (0.10 mL, 0.67 mmol, 1.1 eq.) under inert atmosphere and cooled in an ice bath. MsCl (0.10 mL, 1.34 mmol, 2.2 eq.) was added dropwise, the reaction was allowed to warm to ambient temperature and stirred overnight. Dichloromethane was removed under reduced pressure, EtOAc (15 mL) was added and the organic layer was washed with water (3×15 mL) after phases were separated. The aqueous layer was reextracted with EtOAc (3 ×10 mL) and the combined organic layers were dried over Na_2SO_4. After evaporation and chromatographic purification (Cy/EtOAc, 7/3, R_f = 0.58) **76** (170 mg, 0.70 mmol, quant.) was yielded as orange liquid.

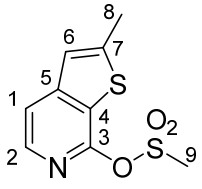
Analytics:

Chemical Formula: $C_9H_9NO_2S$.

Molecular Weight [g mol^{-1}]: 243.30.

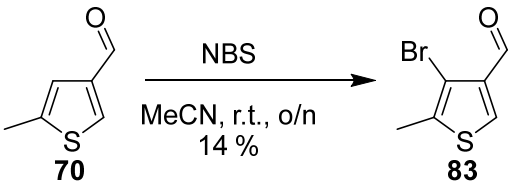
^{1}H-NMR (CDCl$_3$, 400 MHz): δ [ppm] = 8.12 (d; $^3J_{H\text{-}2/H\text{-}1}$ = 5.4 Hz; 1H; H-2), 7.47 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 5.4 Hz; 1H; H-1), 7.11–7.02 (m; 1H; H-6), 3.63 (s; 3H; H-9), 2.65 (d; $^3J_{H\text{-}8/H\text{-}6}$ = 1.1 Hz; 3H; H-8).

HR MS (EI): calculated for $[M]^{\bullet+}$ ($C_9H_9NO_3S_2^{\bullet+}$): 243.0018; found: *m*/*z* = 243.0020.



76

5.2.7.28 Synthesis of 4-bromo-5-methyl-thiophene-3-carbaldehyde (83)



70 **83**

Experimental procedure:

Bromination was achieved according to Hu *et al.*[423]

Aldehyde **70** (1.97 g, mixture with 2-methyl isomer **71** and thiophene-3-carboxaldehyde **18**, 15.6 mmol, 1.0 eq.) and NBS (3.06 g, 17.2 mmol, 1.1 eq.) were dissolved in MeCN (70 mL) and stirred overnight at room temperature. After *in vacuo* concentration the crude mixture was chromatographed (Cy/EtOAc, 5/1, R_f = 0.64) to give **83** (453 mg, 2.2 mmol, 14 % of total starting material) as an orange oil with crystallization tendency.

Analytics:

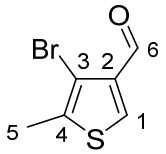
Chemical Formula: C_6H_5BrOS.

Molecular Weight [g mol^{-1}]: 205.07.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 9.89 (s; 1H; H-6), 7.31 (s; 1H; H-1), 2.71 (s; 3H; H-5).

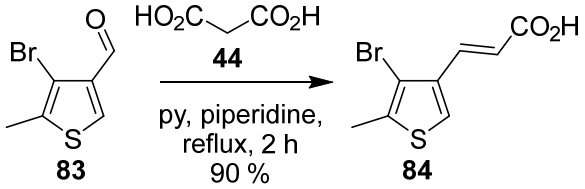
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 183.2 (C-6), 153.0 (C-2), 137.8 (C-4), 129.5 (C-1), 109.7 (C-3), 13.6 (C-5).

HR MS (EI): calculated for $[M-H]^{\bullet+}$ ($C_6H_4BrOS^{\bullet+}$): 202.9161; found: *m*/*z* = 202.9167.



83

5.2.7.29 Synthesis of (*2E*)-3-(4-bromo-5-methyl-thiophen-3-yl)-acrylic acid (84)



83 **84**

Experimental procedure:

Knoevenagel condensation to acrylic acid **84** was performed according to New *et al.*[306]

Aldehyde **83** (220 mg, 1.07 mmol, 1.0 eq.), malonic acid (**44**, 112 mg, 1.07 mmol, 1.0 eq.), and piperidine (0.05 mL, 46 mmol, 0.5 eq.) were dissolved in pyridine (5 mL) and refluxed for 2 h. After cooling to ambient temperatures, the mixture was carefully poured in ice-cold HCl conc. (10 mL). The resulting suspension was extracted with EtOAc (3×15 mL) and evaporated. Acrylic acid **84** (238 mg, 0.96 mmol, 90 %) was obtained as beige powder.

Analytics:

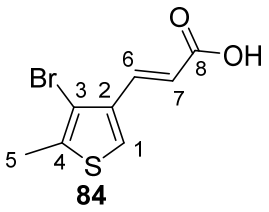
Chemical Formula: $C_8H_7BrO_2S$.

Molecular Weight [g mol^{-1}]: 247.11.

^{1}H-NMR (CD_3OD, 500 MHz): δ [ppm] = 7.61 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 15.8 Hz; 1H; H-6), 7.30 (s; 1H; H-1), 6.20 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 15.7 Hz; 1H; H-7), 2.47 (s; 3H; H-5).

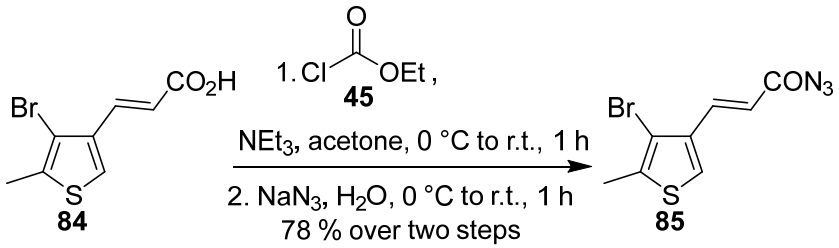
^{13}C-NMR (CD_3OD, 101 MHz): δ [ppm] = 170.5 (C-8), 145.1 (C-4), 136.4 (C-6), 135.2 (C-2), 129.3 (C-1), 119.2 (C-7), 110.3 (C-3), 13.0 (C-5).

HR MS (ESI$^-$): calculated for $[M-H]^-$ ($C_8H_6BrO_2S^-$): 244.9277; found: *m*/*z* = 244.9279.



84

5.2.7.30 Synthesis of (*2E*)-3-(4-bromo-5-methyl-thiophen-3-yl)-acryloyl azide (85)



Experimental procedure:

Acid azide **85** was prepared as described in the literature.[307] Further conversion to the corresponding pyridone derivative in analogy to compound **20** was not achieved.

Acrylic acid **84** (230 mg, 0.93 mmol, 1.0 eq.) and NEt_3 (0.14 mL, 1.02 mmol, 1.1 eq.) were dissolved in acetone (6 mL) and cooled in an ice bath. Ethyl chloroformate (**45**, 0.10 mL, 1.02 mmol, 1.1 eq.) was added dropwise and the mixture was stirred for 1 h while slowly warming to ambient temperatures. Subsequently, a solution of NaN_3 (91 mg, 1.40 mmol, 1.5 eq.) in water (2 mL) was added at 0 °C. After 1 h stirring with removed ice bath the reaction mixture was poured into water (15 mL). The resulting precipitate was collected by filtration, washed with water and dried by lyophilization overnight. Acryoyl azide **85** was obtained as beige solid (198 mg, 0.73 mmol, 78 %).

Analytics:

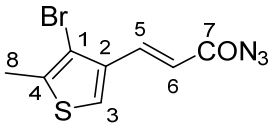
Chemical Formula: $C_8H_6BrN_3OS$.

Molecular Weight [g mol^{-1}]: 272.12.

^{1}H-NMR (CD_3OD, 700 MHz): δ [ppm] = 7.69 (d; $^3J_{H\text{-}5/H\text{-}6}$ = 15.6 Hz; 1H; H-5), 7.34 (s; 1H; H-3), 6.26 (d; $^3J_{H\text{-}6/H\text{-}5}$ = 15.6 Hz; 1H; H-6), 2.50 (s; 3H; H-8).

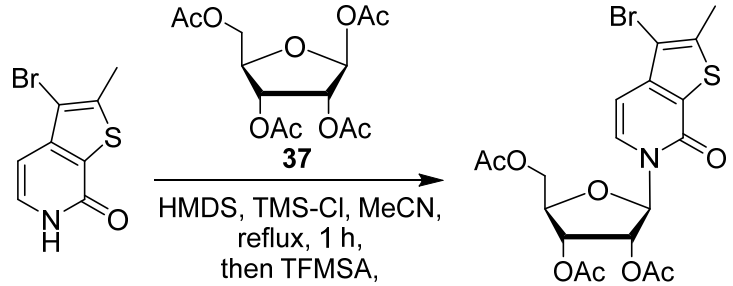
^{13}C-NMR (CD_3OD, 176 MHz): δ [ppm] = 173.7 (C-7), 147.4 (C-4), 137.8 (C-5), 135.0 (C-2), 129.3 (C-3), 119.8 (C-6), 110.7 (C-1), 28.0 (C-8).

HR MS (EI): calculated for $[M]^{\bullet+}$ ($C_8H_6BrN_3OS^{\bullet+}$): 270.9409; found: *m*/*z* = 270.9417.



85

5.2.7.31 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)-3-bromo-2-methyl-6*H*-thieno-[2,3-*c*]pyridin-7-one (104)



Experimental procedure:

The following reaction was performed according to an adapted description by Vorbrüggen and Bennua.[317]

Pyridone derivative **82** (125 mg, 0.51 mmol, 1.0 eq.) was dissolved in dry MeCN (12 mL) under an atmosphere of argon. HMDS (0.11 mL, 0.56 mmol, 1.1 eq.) and TMS-Cl (0.08 mL, 0.61 mmol, 1.2 eq.) were added consecutively and the mixture was heated to reflux until the solution became clear (~1 h). After the addition of peracylated β-D-ribofuranose (**37**, 163 mg, 0.51 mmol, 1.0 eq.) and TFMSA (0.05 mL, 0.61 mmol, 1.2 eq.) at ambient temperature the reaction mixture was stirred at room temperature for 2 d. The reaction was quenched with saturated aqueous $NaHCO_3$ solution (15 mL) and vigorous stirring for 15 min. Filtration facilitated reisolation of unreacted pyridone **82** (56 mg, 0.23 mmol, 45 %). Extraction of the aqueous layer with CH_2Cl_2 (3×20 mL), drying over Na_2SO_4 and removal of the solvents under reduced pressure gave the crude product which was purified by preparative HPLC chromatography (50→80 % MeCN/0.1 % formic acid in 6 min, flow rate 40 mL min^{-1}) yielding **104** (16 mg, 0.03 mmol, 11 % calculated from reactive fraction of **82**) as a light brown oil of very high viscosity.

Analytics (specific assignment given wherever possible):

Chemical Formula: $C_{19}H_{20}BrNO_8S$.

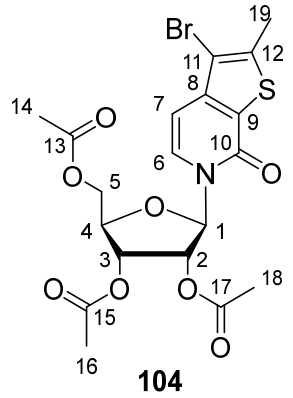
Molecular Weight [g mol^{-1}]: 502.33.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 7.44 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.4 Hz; 1H; H-6), 6.62 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.4 Hz; 1H; H-7), 6.42 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 4.6 Hz; 1H; H-1), 5.49–5.33 (m; 2H; H-5),

4.46–4.29 (m; 3H; H-2; H-3; H-4), 2.54 (s; 3H; H-19), 2.17 (s; 3H; H-14), 2.11 (s; 3H; H-16/18), 2.09 (s; 3H; H-18/16).

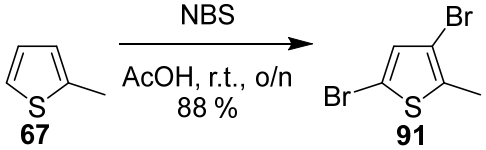
^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 170.4 (C-13), 169.7 (C-15/17), 169.6 (C-17/15), 157.3 (C-10), 144.8 (C-9), 144.5 (C-11), 128.8 (C-6), 126.7 (C-8), 106.9 (C-12), 102.9 (C-7), 88.0 (C-1), 79.7 (C-4), 74.0 (C-2), 70.1 (C-3), 63.1 (C-5), 21.0 (C-14), 20.6 (C-16/18), 20.6 (C-18/16), 16.0 (C-19).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{19}H_{20}BrNO_8SNa^+$): 523.9985; found: *m*/*z* = 523.9978.



104

5.2.7.32 Synthesis of 3,5-dibromo-2-methylthiophene (91)



67 **91**

Experimental procedure:

Bromination took place according to M. Irie *et al*.[424]

2-Methylthiophene (**67**, 5.00 g, 51 mmol, 1.0 eq.) was dissolved in glacial acetic acid (250 mL), NBS (21.8 g, 122 mmol, 2.4 eq.) was added sequentially. After stirring overnight at room temperature the solution was poured on ice-cold water (400 mL), extracted with cyclohexane (3×400 mL), washed with saturated aqueous $NaHCO_3$ solution and brine (400 mL each), dried over Na_2SO_4, and evaporated. Purification over a short plug of silica (Cy) gave dibrominated thiophene **91** (11.5 g, 45 mmol, 88 %) as colorless to faint yellow liquid.

Analytics:

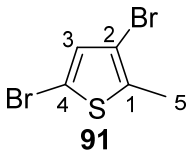
Chemical Formula: $C_5H_4Br_2S$.

Molecular Weight [g mol^{-1}]: 255.96.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 6.75 (s; 1H; H-3), 2.24 (s; 3H; H-5).

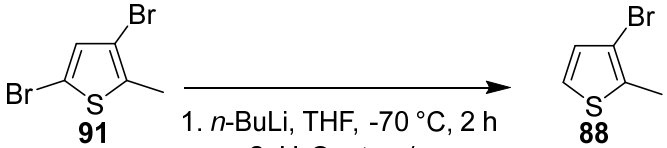
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 136.1 (C-1), 132.1 (C-3), 108.8 (C-2/4), 108.6 (C-4/2), 15.0 (C-5).

MS (EI): calculated for $[M-H]^{\bullet+}$ ($C_5H_3Br_2S^{\bullet+}$): 254.8; found: *m*/*z* = 254.8.



91

5.2.7.33 Synthesis of 3-bromo-2-methylthiophene (88)



91 **88**

Experimental procedure:

The reaction was adapted from M. Irie *et al*.[424]

Dibrominated thiophene **91** (11.3 g, 44 mmol, 1.0 eq.) was dissolved in dry THF (50 mL) and cooled to -70 °C. A solution of *n*-BuLi (2.5 M in hexane, 18 mL, 44 mmol, 1.0 eq.) was added dropwise and the reaction was stirred at -70 °C for 2 h. Upon addition of water (25 mL) the quenched reaction was allowed to warm to room temperature and stirred overnight. After extraction with Et_2O (3×50 mL), drying over Na_2SO_4 and *in vacuo* concentration the crude product was purified by passing through a plug of silca gel (Cy). Thiophene **88** (8.13 g containing residual Cy and CH_2Cl_2, quant.) was obtained as yellow liquid. Due to high volatility of the product **88** residual solvents were not completely removed. **88** was stored at -20 °C.

Analytics:

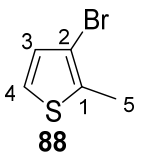
Chemical Formula: C_5H_5BrS.

Molecular Weight [g mol^{-1}]: 177.06.

^{1}H-NMR (CDCl$_3$, 400 MHz): δ [ppm] = 7.07 (d; $^3J_{H\text{-}4/H\text{-}3}$ = 5.4 Hz; 1H; H-4/3), 6.90 (d; $^3J_{H\text{-}3/H\text{-}4}$ = 5.4 Hz; 1H; H-3/4), 2.41 (s; 3H; H-5).

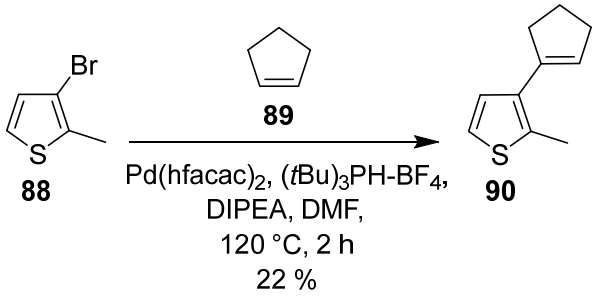
^{13}C-NMR (CDCl$_3$, 101 MHz): δ [ppm] = 134.2 (C-1), 130.0 (C-3), 122.9 (C-4), 109.5 (C-2), 14.7 (C-5).

MS (EI): calculated for $[2M]^{\bullet+}$ ($C_{10}H_{10}Br_2S_2^{\bullet+}$): 351.9; found: *m*/*z* = 351.8.



88

5.2.7.34 Synthesis of 3-cyclopent-1-enyl-2-methyl-thiophene (90)



Experimental procedure:

Cyclic Heck reaction protocol was taken from X. Wu and J. Zhou.[402]

Under inert atmosphere $Pd(hfacac)_2$ (83 mg, 0.16 mmol, 4 mol%) and $(t\text{Bu}_3)\text{P}\cdot\text{BF}_4$ (93 mg, 0.32 mmol, 8 mol%) were dissolved in dry DMF (15 mL), degassed with a stream of argon and stirred for 45 min at room temperature. Then a degassed solution of **88** (700 mg, 4 mmol, 1.0 eq) in dry DMF (5 mL), cyclopentene (**89**, 545 mg, 8.0 mmol, 2.0 eq.), and DIPEA (1.0 mL, 6.0 mmol, 1.5 eq.) were added and the reaction was heated to 120 °C for 2 h. Removal of solvent *in vacuo* and column chromatography (Cy/EtOAc, 10/1, R_f = 0.65) gave **90** (140 mg, 0.9 mmol, 22 %) as a colorless to yellowish liquid.

Analytics:

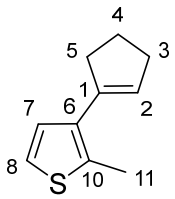
Chemical Formula: $C_{10}H_{12}S$.

Molecular Weight [g mol^{-1}]: 164.27.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 7.03–6.95 (m; 2H; H-7; H-8), 5.86 (m; 1H; H-2), 2.78–2.65 (m; 2H; H-5/3), 2.57–2.47 (m; 5H; H-11; H3/5), 2.04–1.90 (m; 2H; H-4).

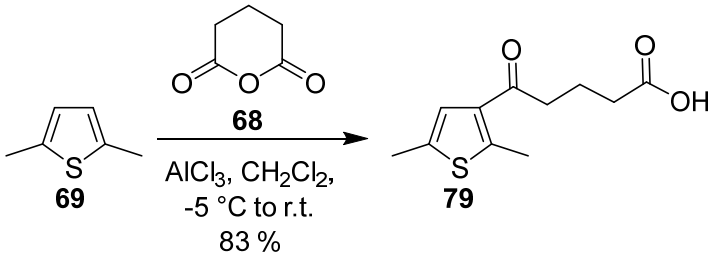
^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 138.4 (C-10), 134.7 (C-1), 133.9 (C-6), 128.3 (C-7), 127.3 (C-2), 120.7 (C-8), 36.0 (C-5/3), 33.5 (C-3/5), 23.5 (C-4), 15.3 (C-11).

HR MS (ESI$^+$): calculated for $[M+H]^+$ ($C_{10}H_{12}SH^+$): 165.0732; found: m/z = 165.0737.



90

5.2.7.35 Synthesis of 5-(2,5-dimethyl-thiophen-3-yl)-5-oxo-pentanoic acid (79)



Experimental procedure:

Under an atmosphere of argon 2,5-dimethylthiophene (**69**, 1.0 g, 8.9 mmol, 1.0 eq.) and glutaric anhydride (**68**, 1.22 g, 10.7 mmol, 1.2 eq) were dissolved in dry CH_2Cl_2 (15 mL) and cooled in an ice/salt bath to ca. -5 °C. $AlCl_3$ (2.97 g, 22.3 mmol, 2.5 eq.) was added spatulawise. The resulting purplish solution was stirred under ice cooling for 2 h, was allowed to reach room temperature and stirred overnight. Quenched by careful addition of ice the reaction mixture was extraced with $CHCl_3$ (3×20 mL), washed with brine and dried over Na_2SO_4 to give the crude product which was purified by column chromatography (Cy/EtOAc + 0.1 % formic acid, 1/1, R_f = 0.48) yielding acid **79** (1.67 g, 7.4 mmol, 83 %).

Analytics (specific assignment given wherever possible):

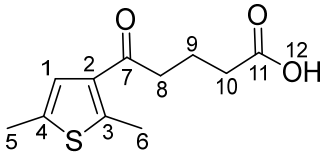
Chemical Formula: $C_{11}H_{14}O_3S$.

Molecular Weight [g mol^{-1}]: 226.29.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 9.68 (s; 1H; H-12), 6.99 (q; $^3J_{H\text{-}1/H\text{-}5}$ = 1.2 Hz; 1H; H-1), 2.87 (t; $^3J_{H\text{-}8/H\text{-}9}$ = 7.1 Hz; 2H; H-8), 2.65 (s; 3H; H-6), 2.47 (t; $^3J_{H\text{-}10/H\text{-}9}$ = 7.2 Hz; 1H; H-10), 2.42–2.37 (m; 3H; H-5), 2.02 (pseudo p; $^3J_{H\text{-}9/H\text{-}8} = {}^3J_{H\text{-}9/H\text{-}10}$ = 7.0 Hz; 2H; H-9).

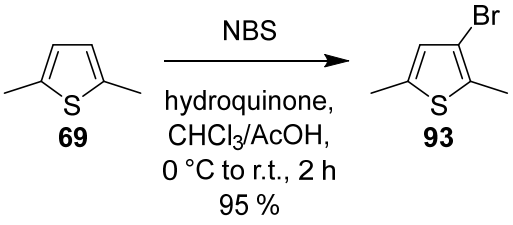
^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 195.6 (C-7), 179.5 (C-11), 147.7 (C-3/2/4), 135.4 (C-2/4/3), 135.3 (C-4/2/3), 126.0 (C-1), 40.5 (C-8), 33.3 (C-10), 19.1 (C-9), 16.2 (C-6), 15.1 (C-5).

HR MS (ESI$^-$): calculated for $[M\text{-}H]^-$ ($C_{11}H_{13}O_3S^-$): 225.0591; found: *m*/*z* = 225.0579.



79

5.2.7.36 Synthesis of 3-bromo-2,5-dimethyl-thiophene (93)



Experimental procedure:

Bromination of **69** was adapted from Rossi *et al*.[425]

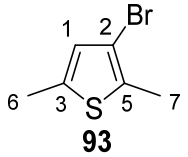
Dimethylthiophene **69** (7.0 g, 62 mmol, 1.0 eq.) was dissolved in $CHCl_3$ (50 mL) and MeCN (50 mL) and cooled in an ice bath. Hydroquinone (14 mg, 0.12 mmol, 0.2 mol%) and NBS (11.0 g, 62 mmol, 1.0 eq) were added and the mixture was allowed to reach room temperature and stirred for 2 h. The reaction was diluted with water (30 mL) and extracted with CH_2Cl_2 (3×100 mL). Combined organic layers were washed with saturated aqueous $NaHCO_3$ solution (50 mL) and water (50 mL), dried over Na_2SO_4, and evaporated to give **93** (11.3 g, 59 mmol, 95 %) as brown liquid which crystallizes over time in sufficient purity for further reaction steps.

Analytics (specific assignment given wherever possible):

Chemical Formula: C_6H_7BrS.

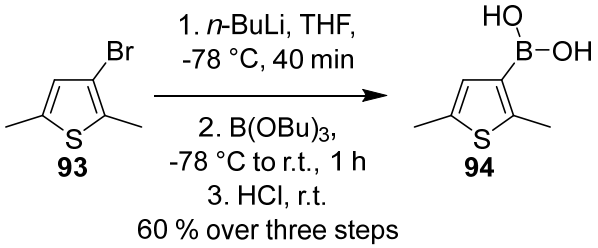
Molecular Weight [g mol^{-1}]: 191.09.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 6.56 (m; 1H; H-1), 2.41 (s; 3H; H-6/7), 2.34 (s; 3H; H-7/6).



93

5.2.7.37 Synthesis of *B*-(2,5-dimethyl-3-thienyl)-boronic acid (94)



Experimental procedure:

Synthesis of boronic acid species **94** was performed according to a description by Mörtel *et al*.[290]

Under inert atmosphere thiophene **93** (1.0 g, 5.2 mmol, 1.0 eq.) was dissolved in dry THF (15 mL) and cooled to -78 °C. BuLi (2.5 M in hexane, 2.3 mL, 5.8 mmol, 1.1 eq.) was added

dropwise and the reaction was stirred at -78 °C for 40 min. After the addition of $B(OBu)_3$ (1.7 mL, 6.3 mmol, 1.2 eq.) the reaction was allowed to warm to room temperature within 1 h. $HCl_{aq.}$ (2 M, 15 mL) was added and the aqueous layer was extracted with Et_2O (3×20 mL) and the combined organic layers were extensively washed with H_2O before extracting the product into $NaOH_{aq.}$ (2 M, 20 mL) which was then acidified with conc. HCl until a white precipitate formed. The suspension was stored at 8 °C overnight and product **94** (482 mg, 3.1 mmol, 60 %) was obtained as white powder after collection by filtration, extensive water washings and overnight lyophilization.

Analytics (specific assignment given wherever possible):

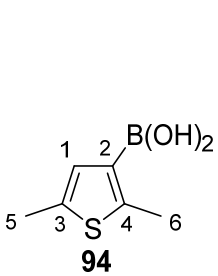
Chemical Formula: $C_6H_9BO_2S$.

Molecular Weight [g mol^{-1}]: 156.01.

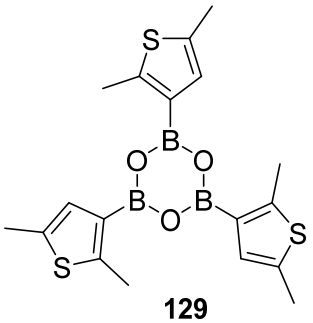
^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 7.07 (q; $^4J_{H\text{-}1/H\text{-}5}$ = 1.3 Hz; 1H; H-1), 2.81 (s; 3H; H-6), 2.45 (d; $^4J_{H\text{-}5/H\text{-}1}$ = 1.0 Hz; 3H; H-5).

^{13}C-NMR ($CDCl_3$, 126 MHz): δ [ppm] = 153.8 (C-2), 136.0 (C-3/4), 131.6 (C-1), 16.2 (C-6), 15.0 (C-5).

MS (EI): calculated for trimer **129** $[3M\text{-}3H_2O]^{\cdot+}$ ($C_{18}H_{21}B_3O_3S_3^{\cdot+}$): 414.1; found: *m*/*z* = 414.1.

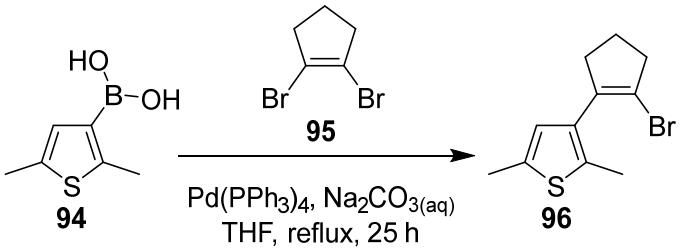


94



129

5.2.7.38 Synthesis of 3-(2-bromo-cyclopent-1-enyl)-2,5-dimethyl-thiophene (96)



94 **96**

Experimental procedure:

The experimental Suzuki coupling procedure was adapted from M. Singer and A. Jäschke[295] Under an atmosphere of argon thiophene derivative **94** (760 mg, 4.9 mmol, 1.1 eq.), 1,2-dibromocyclopentene (**95**, 2.3 mL, 4.4 mmol, 1.0 eq.), and $Pd(PPh_3)_4$ (508 mg, 0.4 mmol, 10 mol%) were dissolved in THF (25 mL) and a degassed aqueous solution of Na_2CO_3 (20 wt%, 1.9 M, 6 mL, 11.4 mmol, 2.6 eq.) was added before refluxing for 20 h. TLC showed

incomplete conversion of starting material; therefore, another spatula tip of catalyst was added and reflux was continued for 4 h. After cooling to room temperature, the reaction was diluted with water (40 mL) and extracted with EtOAc (3x40 mL), dried over Na_2SO_4, and evaporated. Column chromatography (Cy, R_f = 0.70) afforded compound **96** (990 mg, 3.9 mmol, 89 %) as a colorless liquid, which is possibly acid-labile and prone to degradation by elevated temperatures, air, and/or moisture and should rapidly be employed for next synthetic steps or stored under inert gas at -20 °C or lower.

Analytics (specific assignment given wherever possible):

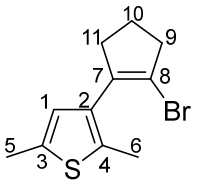
Chemical Formula: $C_{11}H_{13}BrS$.

Molecular Weight [g mol^{-1}]: 257.19.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 6.55 (q; $^4J_{H\text{-}1/H\text{-}5}$ = 1.1 Hz; 1H; H-1), 2.82–2.75 (m; 2H; H-11/9), 2.63–2.55 (m; 2H; H-9/11), 2.40 (s; 3H; H-5/6), 2.33 (s; 3H; H-6/5), 2.09–1.99 (m; 2H; H-10).

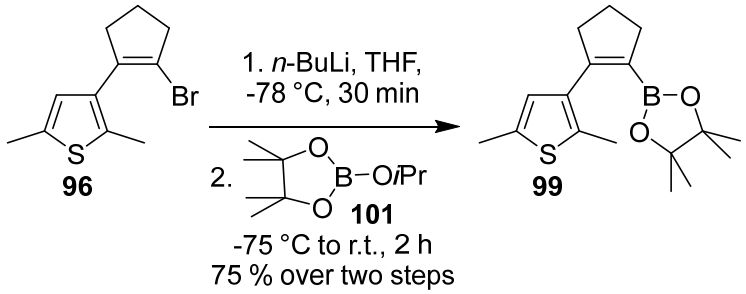
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 140.7 (C-3/4), 135.1 (C-4/3), 133.0 (C-1), 126.1 (C-2), 125.8 (C-7), 61.4 (C-8), 38.5 (C-11/9), 36.4 (C-9/11), 31.4 (C-10), 15.1 (C-5/6), 15.0 (C-6/5).

HR MS (EI): calculated for $[M]^{\bullet+}$ ($C_{11}H_{13}BrS^{\bullet+}$): 255.9916; found: *m*/*z* = 255.9918.



96

5.2.7.39 Synthesis of 2-(2,5-dimethyl-3-thienyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (99)



Experimental procedure:

Formation of boronic acid pinacol ester derivative **99** followed a description by M. Singer and A. Jäschke.[296]

96 (400 mg, 1.6 mmol, 1.0 eq.) was dissolved in dry THF (20 mL) under an atmosphere of argon and cooled in an iPrOH/dry ice bath. An *n*-BuLi solution (2.5 M in hexane, 0.7 mL,

1.7 mL, 1.1 eq.) was added dropwise and stirred for 30 min before the addition of 2-isopropylboronic acid pinacol ester (**101**, 0.4 mL, 1.9 mmol, 1.2 eq.). After 2 h, the reaction was quenched with methanol (3 mL) and the mixture was diluted with EtOAc (40 mL). The organic layer was washed with water (3×40 mL), dried over Na_2SO_4, and evaporated. Purification via column chromatography (Cy/CH_2Cl_2, 1/1, R_f = 0.65) afforded **99** (364 mg, 1.2 mmol, 75 %) as a slightly yellow oil.

Analytics (specific assignment given wherever possible):

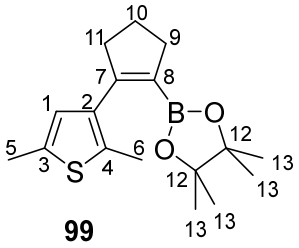
Chemical Formula: $C_{17}H_{25}BO_2S$.

Molecular Weight [g mol^{-1}]: 304.26.

^{1}H-NMR ($CDCl_3$, 400 MHz): δ [ppm] = 6.52 (q; $^4J_{H\text{-}1/H\text{-}5}$ = 1.1 Hz; 1H, H-1), 2.71–2.48 (m; 4H; H-9; H-11), 1.89 (pseudo p; $^3J_{H\text{-}10/H\text{-}11} = {}^3J_{H\text{-}10/H\text{-}19}$ = 7.5 Hz; 2H, H-10), 1.20 (s; 12H; H-13).

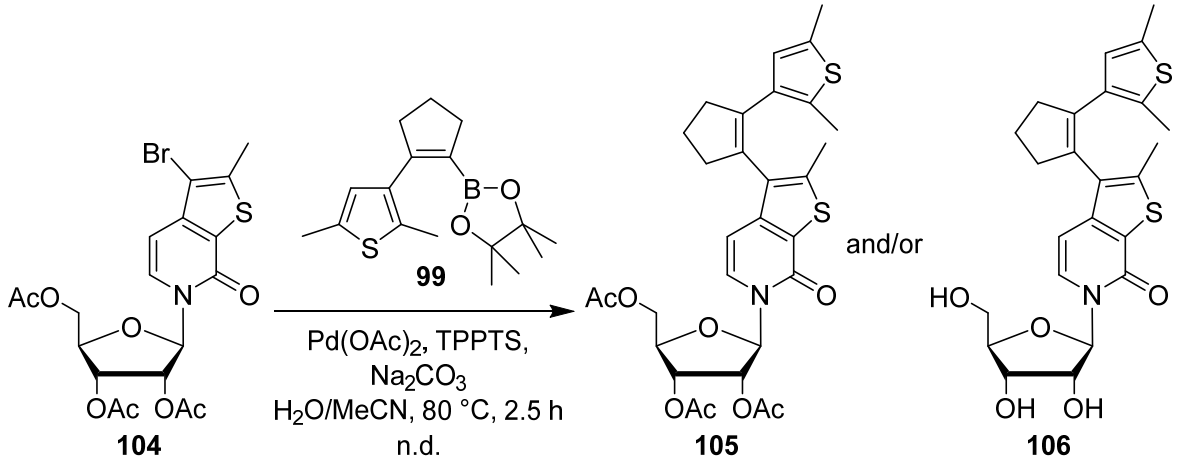
^{13}C-NMR ($CDCl_3$, 101 MHz): δ [ppm] = 152.3 (C-7), 136.3 (C-2), 134.1 (C-3), 132.2 (C-4), 127.3 (C-1), 127.0 (C-8), 82.9 (C-12), 40.3 (C-9/11), 37.2 (C-11/9), 24.9 (C-13), 24.7 (C-10), 15.2 (C-5), 14.5 (C-6).

HR MS (EI): calculated for $[M]^{\bullet+}$ ($C_{17}H_{25}BO_2S^{\bullet+}$): 303.1669; found: *m*/*z* = 303.1698.



99

5.2.7.40 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)-3-[2-(2,5-dimethyl-thiophen-3-yl)-cyclopent-1-enyl]-2-methyl-6*H*-thieno[2,3-*c*]pyridin-7-one (105) and 1′-(β-D-ribofuranosyl)-3-[2-(2,5-dimethyl-thiophen-3-yl)-cyclopent-1-enyl]-2-methyl-6*H*-thieno[2,3-*c*]pyridin-7-one (106)



Experimental procedure:

Suzuki type cross coupling in aqueous media was performed adapting a procedure by G. Hervé and C. Len.[426]

Under an atmosphere of argon a degassed solution of acyl protected nucleoside **104** (15 mg, $29.9 \cdot 10^{-3}$ mmol, 1.0 eq.) and boronic acid pinacol ester **99** (11 mg, $35.8 \cdot 10^{-3}$ mmol, 1.2 eq.) in H_2O/MeCN (2/1, *v*/*v*, 3.5 mL) was added to $Pd(OAc)_2$ (0.7 mg, $3.0 \cdot 10^{-3}$ mmol, 10 mol%), TPPTS (3.4 mg, $6.0 \cdot 10^{-3}$ mmol, 20 mol%), and Na_2CO_3 (6.6 mg, $62.7 \cdot 10^{-3}$ mmol, 2.1 eq.). The resulting mixture was refluxed for 2.5 h, cooled to ambient temperature, concentrated *in vacuo* and freeze dried. The crude product was purified via preparative HPLC (50→100 % MeCN/0.1 % formic acid in 9 min, flow rate 40 mL min^{-1}) to give the deprotected nucleoside **106** (trace amounts, n.d.) as colorless liquid.

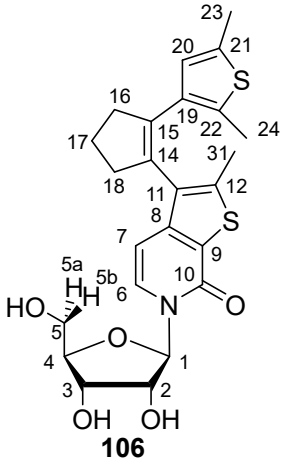
Analytics (specific assignment given wherever possible):

Chemical Formula: $C_{30}H_{33}NO_8S_2$ (**105**), $C_{24}H_{27}NO_5S_2$ (**106**).

Molecular Weight [g mol^{-1}]: 599.71 (**105**), 496.59 (**106**).

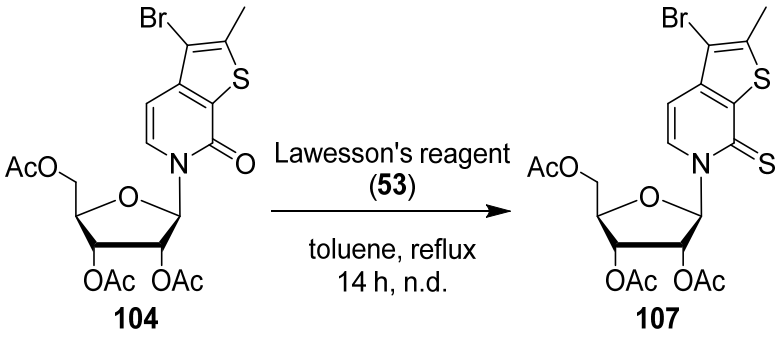
^{1}H-NMR (CD_3OD, 500 MHz): δ [ppm] = 7.82 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.3 Hz; 1H; H-6), 6.59 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.4 Hz; 1H; H-7), 6.40 (q; $^4J_{H\text{-}20/H23}$ = 1.1 Hz; 1H; H-20), 6.26 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 4.0 Hz; 1H; H-1), 4.23–4.18 (m; 2H; H-2, H-3), 4.07 (pseudo dt; $^3J_{H\text{-}4/H\text{-}3}$ = 5.5 Hz; $^3J_{H\text{-}4/H\text{-}5a/b}$ = 2.9 Hz; 1H; H-4), 3.91 (dd; $^2J_{H\text{-}5a/H\text{-}5b}$ = 12.3 Hz; $^3J_{H\text{-}5a/H\text{-}4}$ = 2.7 Hz; 1H; H-5a/5b), 3.79 (dd; $^2J_{H\text{-}5b/H\text{-}5a}$ = 12.3 Hz; $^3J_{H\text{-}5b/H\text{-}4}$ = 3.5 Hz; 1H; H-5b/5a), 2.66 (tt; $^3J_{H\text{-}16/H\text{-}17}$ = 7.8 Hz; $^3J_{H\text{-}16/H\text{-}18}$ = 2.2 Hz; 2H; H-16/18), 2.56 (tt; $^3J_{H\text{-}18/H\text{-}17}$ = 7.5 Hz; $^3J_{H\text{-}18/H\text{-}16}$ = 2.2 Hz; 2H; H-18/16), 2.28 (m; 3H; H-23), 2.17 (s; 3H; H-31), 2.00–1.88 (m; 2H; H-17), 1.79 (s; 3H; H-24).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{24}H_{27}NO_5S_2Na^+$): 496.1223; found: *m*/*z* = 496.1224.



106

5.2.7.41 Synthesis of 1′-(2′,3′,5′-tri-*O*-acetyl-β-D-ribofuranosyl)-3-bromo-2-methyl-6*H*-thieno-[2,3-*c*]pyridin-7-thione (107)



Experimental procedure:

Thienation was achieved adapting a protocol by Li *et al*.[194]

Compound **104** (15 mg, $3.0 \cdot 10^{-2}$ mmol, 1.0 eq.) was dried under high vacuum for several hours. Under an atmosphere of argon Lawesson's reagent (**53**, 18 mg, $4.5 \cdot 10^{-2}$ mmol, 1.5 eq.) was added, both solids were dissolved in dry toluene (3 mL) and refluxed iteratively for a total of 14 h, cooled to ambient temperature, filtered, and evaporated. Purification via preparative HPLC (30→80 % MeCN/0.1 % formic acid in 9 min, flow 40 mL min^{-1}, t_R = 7.7 min, R_f = 0.42 in Cy/EtOAc, 1/1) to give analytically pure **107** (trace amounts, n.d.).

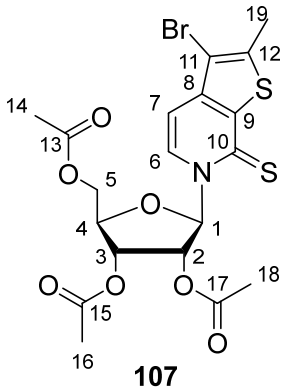
Analytics (specific assignment given wherever possible):

Chemical Formula: $C_{19}H_{20}BrNO_7S_2$.

Molecular Weight [g mol^{-1}]: 518.39.

^{1}H-NMR ($CDCl_3$, 500 MHz): δ [ppm] = 8.10 (d; $^3J_{H\text{-}6/H\text{-}7}$ = 7.3 Hz; 1H; H-6), 7.28 (d; $^3J_{H\text{-}1/H\text{-}2}$ = 2.9 Hz; 1H; H-1), 7.01 (d; $^3J_{H\text{-}7/H\text{-}6}$ = 7.3 Hz; 1H; H-7), 5.55 (dd; $^3J_{H\text{-}2/H\text{-}3}$ = 5.3 Hz; $^3J_{H\text{-}2/H\text{-}1}$ = 2.9 Hz; 1H; H-2), 5.28 (dd; $^3J_{H\text{-}3/H\text{-}4}$ = 7.4 Hz; $^3J_{H\text{-}3/H\text{-}2}$ = 5.4 Hz; 1H; H-3), 4.53 (dt; $^3J_{H\text{-}4/H\text{-}3}$ = 7.4 Hz; $^3J_{H\text{-}4/H\text{-}5}$ = 2.9 Hz; 1H; H-4), 4.47 (t; $^3J_{H\text{-}5/H\text{-}4}$ = 2.8 Hz; 2H; H-5), 2.56 (s; 3H; H-19), 2.22 (s; 3H; H-18/16/14), 2.16 (s; 3H; H-16/18/14), 2.08 (s; 3H; H-14/16/18).

HR MS (ESI$^+$): calculated for $[M+Na]^+$ ($C_{19}H_{20}BrNO_7S_2Na^+$): 539.9757; found: *m*/*z* = 539.9753.



5.3 Biochemical methods

While handling nucleic acids particular caution is advised in order to prevent degradation of valuable sample material by ubiquitous nucleases.[427] Disposable nitrile or latex gloves were worn at all times and sterilized using 80 % EtOH before handling samples. Reaction tubes (0.2 mL, 1.5 mL, 2.0 mL) were autoclaved and stored in sealed glassware. Conical tubes (15 mL, 50 mL, DNA-, DNase-, RNase-free by supplier) were used without further treatment. Pipette tips were autoclaved and stored in sealable boxes.

5.3.1 General biochemical methods

Ultrapure water

For any biochemical application autoclaved and sterile filtered ultrapure water (referred to as ddH_2O) was employed. Ultrapure water was generated from desalted water using a *Barnstead MicroPure* water purification system (*Thermo Fisher Scientific*).

Synthetically synthesized DNA/RNA oligonucleotides

All unmodified sequences used in this study were ordered from *metabion* (Planegg/Steinkirchen, Germany), *biomers.net* or *Ella Biotech* (Martinsried, Germany). DNA templates containing unnatural nucleotides (d**NaM**, d**5SICS**) bases were synthesized by *Ella Biotech*, the corresponding CEP building blocks were purchased from *Berry & Associates Inc.* (Dexter, MI, USA) and provided to *Ella Biotech*. Synthetic nucleic acids were produced, purified, analyzed, and freeze dried by the manufacturers.

Generally, chemically synthesized nucleic acids were dissolved in appropriate volumes of ddH_2O to give 100 µM stock solutions and subsequently stored at -20 °C.

5.3.2 Equipment

Table 6. List of equipment used for biochemical experiments in this study.

Equipment	Model	Brand
96-well plate	Black, low-binding, half area	*Corning*
Agarose gel electrophoresis chamber	n.a.	made in-house
Autoclave	*D65*	*Systec*
Benchtop centrifuge	*Micro 2000*, *Sorvall fresco*	*Hettich, Thermo Fisher*
Block heater	*HBT-1-131*	*HLC BioTech*
C18 cartridges	*Sep-Pak*	*Waters*
Centrifugal filters	*Amicon 3K*	*Merck*

Table 6 (continued)

Conical tubes	15 mL, 50 mL	*Sarstedt*
Electroelution chamber	n.a.	made in-house
Fluorescence spectrometer	*LS 55*	*PerkinElmer*
Freeze drier	*LCG Lyo Chamber Guard*	*Martin Christ*
Gel documentation system	*Gel Doc 2000*, *Genoplex, BioDoc-It™*	*Bio-Rad, VWR*, *UVP*
Handheld UV lamp	*n.a.*	*Herolab*
High precision cell	0.1 cm for T_mK	*Hellma Analytics*
HPLC columns	*Nucleodur, Zorbax, XTerra*	*Macherey-Nagel, Waters*
HPLCs	*Agilent 1100*/*1260 Infinity* (*II*)	*Agilent Technologies*
Mass spectrometer	*HTC esquire*	*Bruker Daltonic*
Cuvettes	*Ultra-Micro*	*Hellma Analytics*
PAGE cask (analytical)	*Mini Protean*	*Bio-Rad*
PAGE cask (preparative)	n.a.	made in-house
pH meter	*FiveEasy Plus*	*Mettler Toledo*
Phosphorimager	*FLA-3000*	*Fujifilm*
Pipette tips	10 µL, 200 µL, 1 mL	*Sarstedt*
Pipette tips 200 µL	*Mµltiflex*	*Carl Roth*
Plate reader	*Ultra*	*Tecan*
Power supply for PAGE	*P25, E132*	*Biometra, Consort*
Reaction tubes	0.2 mL, 1.5 mL, 2.0 mL	*Sarstedt*
Spectrophotometer	*Nanodrop 2000c*	*Thermo Fisher*
Spectropolarimeter	*J-810*	*JASCO*
Spin columns	*G-25 Illustra MicroSpin, NucleoSpin Gel and PCR Clean-up kit, Amicon® Ultra-0.5 Centrifugal Filters 3K*	*GE Healthcare, Macherey Nagel, Merck Millipore*
Syringe filters	Single use, 0.22 µm, sterile	*Labsolute (Th. Geyer)*
Thermocycler	*LifeECO*, *Bioer*	*Biozym, Biometra*
TLC plates	*Silica Gel 60 F254*, 20×20 cm	*Merck*
UV-Vis spectrophotometer	*Cary 100*	*Agilent Technologies*

5.3.3 Reagents and buffers

Table 7. List of reagents used in biochemical experiments.

Reagent	**Supplier**
(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES)	*Applichem*
Acetonitrile	*Thermo Fisher*, *ABCR*, *VWR*
Alkaline phosphatase	*Promega*
Ammoniumperoxodisulfate (APS)	*Merck*
Benzonase nuclease	*Sigma Aldrich, Merck*
Boric acid	*Alfa Aesar*
Bromophenol blue	*Carl Roth*
Dimethylsulfoxide (DMSO)	*Sigma Aldrich*
Dithiothreitol (DTT)	*Sigma Aldrich*
DNase I	*New England Biolabs*
Ethanol (EtOH) absolute	*VWR*
Ethidium bromide (EtBr)	*Carl Roth*
Ethylenediaminetetraacetic acid (EDTA)	*Carl Roth*, *Thermo Fisher*
Ficoll 400	*Sigma Aldrich*
Formamide (FA)	*Fluka*
GeneRuler Ultra Low Range DNA Ladder	*Thermo Fisher*
D-Glucosamine 6-phosphate (GlcN6P, **1**)	*Sigma Aldrich*
Glycerol	*Sigma Aldrich*
Hexafluoroisopropanol (HFIP), LC-MS grade	*Fluka*
High resolution agarose	*Carl Roth*
Inorganic pyrophosphatase (iPP)	*Roche*, *New England Biolabs*
KCl	*Carl Roth*
$MgCl_2 \cdot 6H_2O$	*Alfa Aesar*
Mineral oil	*Sigma Aldrich*
Na_2HPO_4	*Sigma Aldrich*
NaCl	*Carl Roth*
NaH_2PO_4	*Carl Roth*
NH_4OAc	*Carl Roth*
NH_4OAc, LC-MS grade	*Sigma Aldrich*
rNTPs, dNTPs	*Jena Bioscience*
RNasin (or *RNasin plus*)	*Promega, New England Biolabs*
S1 nuclease	*Thermo Fisher*
Sequencing gel DPAGE	*Carl Roth*

Table 7 (continued)

Sequencing gel native PAGE	*Carl Roth*
Snake venom phosphodiesterase I	*Sigma Aldrich, Merck*
SP6 RNA Polymerase	*New England Biolabs*
SYBR Safe DNA stain	*Life Technologies*
T3 RNA Polymerase	*New England Biolabs*
T7 RNA Polymerase	made in-house
Taq DNA polymerase	made in-house
Tetramethylethylenediamine (TEMED)	*Merck*
Tetrazines ^{H}**Tet-BDP** (**57**), ^{Me}**Tet-BDP** (**59**)	*Jena Bioscience* (#CLK-036, #CLK-037)
Urea	*Carl Roth*
Xylene cyanol	*Applichem*

Table 8. List of buffers and their composition used in this study.

Buffer	**Composition**
10× DNase I reaction buffer (HEPES)*	100 mM HEPES, 25 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.6
10× DNase I reaction buffer (tris, *New England Biolabs*)	100 mM tris, 25 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.6
10× *glmS* cleavage buffer*	500 mM HEPES, 2 M KCl, 10 mM $MgCl_2$, pH 7.5
10× Phosphate buffer*	1.45 M NaCl, 100 mM Na_2HPO_4, 100 mM NaH_2PO_4, pH 7
10× S1 nuclease buffer	300 mM NaOAc, 20 mM $Zn(OAc)_2$, 50 % (*v*/*v*) glycerol
10× *Taq* reaction buffer*	100 mM tris, 500 mM KCl, 1 % (*v*/*v*) Triton X-100, pH 9
10× TB*	890 mM tris, 890 mM boric acid, 250 mM NaCl, pH 8
10× TBE*	890 mM tris, 890 mM boric acid, 20 mM EDTA, pH 8
5× Transcription buffer (HEPES)*	200 mM HEPES pH 7.9
5× Transcription buffer (tris)*	200 mM tris pH 7.9
6× DNA loading buffer*	50 mM tris, 15 % (*w*/*v*) Ficoll 400, 6 mM EDTA, 0.25 % (*w*/*v*) bpb, pH 8.3

*made in-house

5.3.4 Two-step assembly PCRs

To generate full-length dsDNA templates for *in vitro* preparation of *glmS* RNAs a two-step protocol was applied, adapting procedures from Shevchuk *et al.*[348]

Step 1

In a first extension approach the reaction mixtures, prepared in 50 or 100 µL scale, contained 1× *Taq* reaction buffer, 2 mM $MgCl_2$, 375 µM each natural dNTP, 0-200 µM d**NaM** TP and d**TPT3** TP, 1 µM DNA antisense templates **glmS_DNA$^{1n/1NaM/1NaM_P1/1n_2}$** and **glmS_DNA$^{3n/3NaM/3NaM_P4a/3NaM_P4R/3NaM_P4RP4a/3n_P4RP4a_HH/3NaM_P4P4a_HH}$**, and 0.5 µM sense template **glmS_DNA$^{2n/2n_2}$** (for *glmS*-Hammerhead constructs additionally 0.5 µM **glmS_DNA^{4n_HH}** was supplemented), as well as 0.05 U µL^{-1} *Taq* DNA polymerase. An exemplary 100 µL protocol is given in Table 9, the PCR conditions for this first step can be found in Table 10.

Table 9. Pipetting scheme for step 1 of assembly PCRs generating dsDNA templates for *glmS* constructs.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**
10× *Taq* reaction buffer	10×	1×	10
$MgCl_2$	100 mM	2 mM	2
dNTP mix	25 mM each	375 µM each	1.5
d**TPT3** TP	10 mM	0 or 100 µM	0 or 1
d**NaM** TP	10 mM	0 or 100 µM	0 or 1
glmS_DNA^1 template	100 µM	1 µM	1
glmS_DNA^2 template	100 µM	0.5 µM	0.5
glmS_DNA^3 template	100 µM	1 µM	1
glmS_DNA^4 template (only for *glmS*-HH)	100 µM	0.5 µM	0.5
Taq DNA polymerase	5 U µL^{-1}	0.05 U µL^{-1}	1
ddH_2O	-	-	ad 100

Table 10. Parameters for step 1 of assembly PCRs generating dsDNA templates for *glmS* constructs; for *glmS*-Hammerhead constructs 47 °C or 57 °C were employed as annealing temperature.

step	***T* [°C]**	***t* [s]**	
initial denaturation	95	120	
denaturation	95	15	5 cycles
annealing	57*	20	
extension	72	60	
hold	4	∞	

Step 2

3 µL of the crude step 1 products were employed in 100 µL amplification reactions containing 1× *Taq* reaction buffer, 2 mM $MgCl_2$, 375 µM each natural dNTP, 0-200 µM d**NaM** TP and d**TPT3** TP, 1 µM forward and reverse primers, and 0.05 U μL^{-1} *Taq* DNA polymerase. Table 11 shows a detailed pipetting scheme, PCR amplification parameters are stated in Table 12. Optionally, another 5 U *Taq* DNA polymerase were added and the program was repeated to maximize yield.

Table 11. Pipetting scheme for step 2 of assembly PCRs generating dsDNA templates for *glmS* constructs.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**
10× *Taq* reaction buffer	10×	1×	10
$MgCl_2$	100 mM	2 mM	2
dNTP mix	25 mM each	375 µM each	1.5
d**TPT3** TP	10 mM	0 or 100 µM	0 or 1
d**NaM** TP	10 mM	0 or 100 µM	0 or 1
DNA template step 1	-	-	3
glmS_Primer_FW$^{1/2}$	100 µM	2 µM	2
glmS_Primer_RV$^{1/2/NaM_4R/HH}$	100 µM	2 µM	2
Taq DNA polymerase	5 U μL^{-1}	0.05 U μL^{-1}	1
ddH_2O	-	-	ad 100

Table 12. Parameters for step 2 of assembly PCRs generating dsDNA templates for *glmS* constructs.

step	***T* [°C]**	***t* [s]**	
initial denaturation	95	120	
denaturation	95	15	
annealing	54	20	20 cycles
extension	72	60	
final extension	72	180	
hold	4	∞	

Alternative PCR program employed with four overlapping DNA templates

The altered two-step assembly PCR program for generation of the *glmS*-Hammerhead constructs **glmS_RNA$^{CP/NO_P4RP4a_HH}$** (section 3.2.2.3) was adapted from Rydzanicz *et al.* [390]

Step 1

The extension step reaction mixture contained 1× *Taq* reaction buffer, 2 mM $MgCl_2$, 200 µM each natural dNTP as well as d**NaM** TP and d**TPT3** TP, 0.4 µM each DNA template (**glmS_DNA^{1n_2}**, **glmS_DNA^{2n_2}**, **glmS_DNA$^{3n_2/3NaM_P4RP4a_HH}$**, **glmS_DNA^{4n_HH}**). *Taq* DNA polymerase was added to obtain 0.025 U µL^{-1}. A 100 µL pipetting scheme and thermal parameters are given in Table 13 and Table 14, respectively.

Table 13. Pipetting scheme for step 1 of alternative four piece assembly PCRs generating dsDNA templates for *glmS*-Hammerhead constructs.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**
10× *Taq* reaction buffer	10×	1×	10
$MgCl_2$	100 mM	2 mM	2
dNTP mix	5 mM each	200 µM each	4
d**TPT3** TP	10 mM	200 µM	2
d**NaM** TP	10 mM	200 µM	2
glmS_DNA^{1n_2}	10 µM	0.4 µM	4
glmS_DNA^{2n_2}	10 µM	0.4 µM	4
glmS_DNA$^{3n_2/3NaM_P4RP4a_HH}$	10 µM	0.4 µM	4
glmS_DNA^{4n_HH}	10 µM	0.4 µM	4
Taq DNA polymerase	1.25 U µL^{-1}	0.025 U µL^{-1}	2
ddH_2O	-	-	ad 100

Table 14. Parameters for the first step of four-piece assembly PCR generating Hammerhead-flanked *glmS* ribozyme constructs.

step	***T* [°C]**	***t* [s]**	
initial denaturation	94	120	
denaturation	94	90	5 cycles
annealing	50	120	
extension	72	180	
final extension	72	300	
hold	4	∞	

Step 2

The amplification PCR included 1 µL of the crude step 1 product mixture (100 µL reaction scale) in 1× *Taq* reaction buffer supplied with 2 mM $MgCl_2$, 20 µM each natural dNTPas well as d**NaM** TP and d**TPT3** TP, 1.5 µM forward and reverse primers, and 0.025 U μL^{-1} *Taq* DNA polymerase. Pipetting scheme and thermal cycling parameters can be found in Table 15 and Table 16.

Table 15. Pipetting scheme for step 2 of alternative four piece assembly PCRs generating dsDNA templates for *glmS*-Hammerhead constructs.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**
10× *Taq* reaction buffer	10×	1×	10
$MgCl_2$	100 mM	2 mM	2
dNTP mix	5 mM each	200 µM each	4
d**TPT3** TP	10 mM	200 µM	2
d**NaM** TP	10 mM	200 µM	2
DNA template step 1	-	-	1
glmS_Primer_FW[2]	100 µM	0.15 µM	1.5
glmS_Primer_RV[HH]	100 µM	0.15 µM	1.5
Taq DNA polymerase	1.25 U μL^{-1}	0.025 U μL^{-1}	2
ddH_2O	-	-	ad 100

Table 16. Parameters for the second step of four-piece assembly PCR generating Hammerhead-flanked *glmS* ribozyme constructs.

step	***T* [°C]**	***t* [s]**	
initial denaturation	94	300	
denaturation	94	30	
annealing	50/54	120	20 cycles
extension	72	180	
final extension	72	300	
hold	4	∞	

5.3.5 PCR from plasmid

dsDNA templates for *in vitro* generation of **Xist^{n}** and **Xist^{CP}** were prepared by PCR amplifying the *Xist* A region (nucleotides 365 to 766 of plasmid pCMV-Xist-PA, *Addgene*, #26760) using forward primer **Xist_Primer_FW** and reverse primer **Xist_Primer_RV$^{n/NaM}$**. PCR amplification was performed using 30 ng plasmid per 100 µL PCR reaction with 0.5 mM forward and reverse primers, 0.4 mM natural dNTPs, 0.2 mM d**TPT3** TP (**22**) and d**NaM** TP (**24**) in 1× *Taq* reaction buffer supplemented with 2 U *Taq* DNA polymerase. Table 17 shows detailed thermal cycling parameters. PCR products were purified by preparative agarose gel electrophoresis and recovered using *NucleoSpin Gel and PCR Clean-up kit* (*Macherey-Nagel*) according to the manufacturers' protocol.

Table 17. PCR parameters for amplifying a dsDNA template of the *Xist* A region.

step	*T* [°C]	*t* [s]	
initial denaturation	95	120	
denaturation	94	30	
annealing	62	40	15 cycles
extension	72	120	
final extension	72	180	
hold	4	∞	

Amino acid sequence of the employed *Taq* DNA polymerase

Thermus aquaticus DNA polymerase (Pol I, according to GenBank[428]: J04639.1):

MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDA
VIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLA
SLAKKAEKEGYEVRILTADKDLYQLLSDRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRAL
TGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAK
VRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGF
VLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPG
DDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLY
REVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLER
VLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRT
GRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLA
HLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQE
LAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREA
AERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLA
KEVMEGVYPLAVPLEVEVGIGEDWLSAKE.

5.3.6 *In vitro* transcription

In vitro transcription reactions were performed in 20 µL, 100 µL, or 200 µL scale. DNA templates for short constructs (*e.g.* **RNA^{CP}**, **RNA^{n1}**, **RNA$^{sc–NO}$**) were added to a final conentration of 3 µM in equal ratio with a corresponding primer and hybridized by annealing at 95 °C for 2-5 min in buffer and $MgCl_2$ followed by slow or snap cooling to ambient temperature. In case of dsDNA originating from PCRs an appropriate volume of crude or purified template was employed (typically 20 µL crude PCR for a 100 µL transcription or 18-20 pmol PCR template for purified dsDNA templates. Transcription reactions contained final concen-trations of 40 mM buffer at pH 7.9 (Tris, in case of *glmS* constructs HEPES), 25 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 2.5 mM canonical triphosphates, 0-2 mM unnatural base triphosphates (r**TPT3** TP (**23**), r**TPT3^{I}** TP (**109**), r**TPT3^{CP}** TP (**34**), r**TPT3^{NO}** TP (**61**), or r**NaM** TP (**25**)), 0.5 U µL^{-1} RNasin, 0 or 2 ng µL^{-1} inorganic pyrophosphatase (iPP), and 2.5–5 U µL^{-1} T7 RNA polymerase, which was added to the mixture last.

In vitro transcriptions were usually incubated at 37 °C for 4 h. For qualitative evaluation of RNA products, reactions were conducted for 2 h, for yield maximization and longer constructs (*glmS* constructs, *Xist* fragment) samples were incubated overnight.

Exemplary recipes for preparation of *in vitro* transcription reactions can be found in Table 18.

Table 18. Pipetting schemes for 20 µL and 100 µL scale *in vitro* transcription reactions.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**	
			Σ = 20	**Σ = 100**
$MgCl_2$	100 mM	25 mM	5	25
Buffer (Tris/HEPES)	200 mM	40 mM	4	20
rNTP mix	25 mM each	2.5 mM	2	10
DTT	100 mM	5 mM	1	5
UB TP	5-10 mM	0-2 mM	0-2	0-20
RNasin	40 U µL^{-1}	0.5-0.6 U µL^{-1}	0.3	1.2
iPP	0.1 U mL^{-1}	0-3·10^{-7} U µL^{-1}	-	0.3
DNA template*	100 µM	3 µM	0.6	3
DNA primer *	100 µM	3 µM	0.6	3
dsDNA*	varying	varying	varying	varying
T7 RNA polymerase	50 U µL^{-1}	2.5-5 U µL^{-1}	1-2	5-10
ddH_2O	-	-	ad 20	ad 100

* either a template and primer pair or dsDNA originating from a preceding PCR was employed.

Amino acid sequence of the employed T7 RNA polymerase

Enterophage T7 RNA Polymerase, Gene 1 (according to GenBank[428]: AY264774.1):

MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAG
EVADNAAAKPLITTLLPKMIARINDWFEEVKAKRGKRPTAFQFLQEIKPEAVAYITIKTTLACLT
SADNTTVQAVASAIGRAIEDEARFGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAFMQVVEA
DMLSKGLLGGEAWSSWHKEDSIHVGVRCIEMLIESTGMVSLHRQNAGVVGQDSETIELAPE
YAEAIATRAGALAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYED
VYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPVEDIPAIEREELPMKPEDIDMNPEA
LTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSMFN
PQGNDMTKGLLTLAKGKPIGKEGYYWLKIHGANCAGVDKVPFPERIKFIEENHENIMACAKS
PLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNCSLPLAFDGSCSGIQHFSAMLRDEVG
GRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAG
QWLAYGVTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMA
KLIWESVSVTVVAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCAVHWVTPDGFPVWQEY
KKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTVVWAHEK
YGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFYDQFADQLHESQLDKMPA
LPAKGNLNLRDILESDFAFA.

5.3.7 DNase digestion

DNA templates in crude transcription reactions were DNase-digested by adding a suitable amount of 10× DNase reaction buffer (in case of *glmS* constructs HEPES-based) and RNase-free DNase I to a final concentration of 0.04 U μL^{-1}. Samples were incubated at 37 °C for 30-45 min and then heated to 95 °C for 2-3 min for enzyme inactivation. A detailed protocol can be found in Table 19.

Table 19. Pipetting scheme for DNase I digestions in crude transcription reactions.

Component	**Stock concentration**	**Final concentration**	***V* [µL]**	
			Σ = 20	**Σ = 100**
DNase buffer	10×	1.1×	2.5	12.5
DNase I	0.2 U μL^{-1}	$4 \cdot 10^{-3}$ U μL^{-1}	0.5	2.5

5.3.8 Nucleoside digestion

Digestion of RNA or nucleotides was adapted from Tolle *et al*.[414] 3 µL of 5× S1 nuclease reaction buffer were added to 200 pmol RNA (purified by gel filtration) in 26.5 µL water. 0.5 µL S1 nuclease (100 U μL^{-1}) were added and the samples were incubated for 60 min at 37 °C, 800 rpm. 3.5 µL 10× alkaline phosphatase buffer and 0.5 µL alkaline phosphatase (1 U μL^{-1}), 0.5 µL snake venom phosphodiesterase I ($5 \cdot 10^{-2}$ U μL^{-1}) and 0.5 µL benzonase nuclease (250 U μL^{-1}) were added. The samples were incubated for 120 min at 37 °C, 800 rpm before enzyme inactivation at 95 °C for 2 min.

In a similar fashion, appropriate mixtures of nucleoside triphosphates (200-1400 pmol per triphosphate) were digested to nucleosides by incubation with 0.5 U alkaline phosphatase in 1× alkaline phosphatase reaction buffer (total volume 35 µL) for 120 min at 37 °C, 800 rpm before enzyme inactivation at 95 °C for 2 min.

Samples were centrifuged 3 min at 12000×g and 10–20 µL of the supernatant were subjected to HPLC analysis on a *Synergi Fusion* 2×50 mm, 4 µm reversed phase column (*Phenomenex*). As mobile phase (A) an aqueous solution of 10 mM NEt_3 and 100 mM hexafluoroisopropanol was used, solvent B was acetonitrile. The separation run (flow 0.3 mL min^{-1}) implied 3 min 100 % A, followed by a gradient to 20 % B in 17 min and further increasing to 70 % B within 20 min. Exemplary protocols for nucleoside digestions are given in Table 20 and Table 21.

Table 20. Pipetting scheme for RNA digestion to nucleosides.

Component	Stock conc.	*V* [µL]	Final conc.
RNA	50 µM	4.0	6.67 µM
S1 nuclease reaction buffer	10×	3.0	1×
S1 nuclease	1000 U μL^{-1}	0.5	16.7 U μL^{-1}
Alkaline phosphatase buffer	10×	3.5	1×
Alkaline phosphatase	1 U μL^{-1}	0.5	$1.4 \cdot 10^{-2}$ U μL^{-1}
Snake venom phosphodiesterase I	$5 \cdot 10^{-3}$ U μL^{-1}	0.5	$7.0 \cdot 10^{-2}$ mU μL^{-1}
Benzonase nuclease	250 U μL^{-1}	0.5	3.5 U μL^{-1}
ddH_2O	-	22.5	-

Table 21. Exemplary pipetting scheme for nucleoside digestion from a triphosphate mix corresponding to relative amounts in **RNA^{n1}**/**RNA^{TPT3}**/**RNA^{NaM}**.

Component	Stock conc.	Volume [µL]	Final conc.
r**A** TP (**118**)	200 µM	4	23 µM
r**C** TP (**119**)	200 µM	3	17 µM
r**G** TP (**120**)	200 µM	7	40 µM
r**U** TP (**121**)	200 µM	2/2/3	17 µM
rTPT3 TP (**23**)/r**NaM** TP (**25**)/ctrl.	200 µM	1/1/0	6 µM
Alkaline phosphatase buffer	10×	3.5	1×
Alkaline phosphatase	1 U µL^{-1}	0.5	1.4·10^{-2} U µL^{-1}
ddH_2O	-	13/13/14	-

Table 22. Retention times (t_R) of ribonucleosides obtained from HPLC measurements (see above for gradient details).

	r**A** (**122**)	r**C** (**123**)	r**G** (**124**)	r**U** (**125**)	r**TPT3** (**117**)	r**NaM** (**126**)
t_R [min]	12.17	2.36	9.33	3.48	20.76	25.82
	12.15	2.35	9.32	3.44	20.72	25.80
	12.14	2.35	9.22	3.45	20.69	25.79
	12.13	2.34	9.28	3.45	-	25.77
	12.20	2.37	9.41	3.49	-	25.77
	12.17	2.36	9.35	3.49	-	-
	12.19	2.36	9.35	3.45	-	-
	12.21	2.38	9.43	3.51	-	-
	12.21	2.38	9.42	3.51	-	-
	12.33	2.42	9.60	3.57	-	-
Mean [min]	12.19	2.37	9.37	3.48	20.72	25.79
SD [min]	0.06	0.02	0.10	0.04	0.04	0.02

5.3.9 IEDDA click reactions on r**TPT3^{CP}**-modified RNAs

Fluorescent labeling

r**TPT3^{CP}**(*i.e.* cyclopropene-)-modified RNAs were incubated with an excess (3–7-fold) of tetrazine-Oregon Green conjugate **Tet-OG** (**28**) or tetrazine-BODIPY conjugates **$^{H/Me}$Tet-BDP** (**57**, **59**) as a 500 µM or 1 mM solution in DMSO allowing the cycloaddition reaction to proceed at room temperature in the dark for at least 1 h.

Biotin-streptavidin labeling (EMSA)

Click reactions of *glmS* constructs with tetrazine-biotin derivative **58** were performed with 5 µL crude, DNase-digested RNA and 2 µL **58** (500 µM in DMSO) for 1 h at room temperature. Subsequently, the samples were diluted with ddH_2O (20 µL), gel-filtrated, and freeze-dried. Redissolved in ddH_2O (2 µL), the biotinylated RNAs were incubated with 1 µL streptavidin solution (1 mg mL^{-1}, 15.5 U mL^{-1}) at r.t. for 1 h before gel electrophoretic analysis.

5.3.10 Gel electrophoresis

Analytical an preparative separation of oligonucleotides was performed with polyacrylamide (PA) or agarose gel electrophoresis. As internal reference *GeneRuler Ultra Low Range DNA Ladder* or *GeneRuler 100 bp DNA Ladder* were employed (5–7 µL), depending on the size of nucleic acid to be analyzed.

Polyacrylamide gel electrophoresis (PAGE)

For analytical (ca. 5 mL gel volume, gel size 83×73×0.75 mm) and preparative (ca. 50 mL gel volume, gel size 160×175×1.5 mm) denaturing PAGE (DPAGE) separation, a solution of formamide/8.3 M urea (95/5, *v*/*v*) was used as loading buffer in equal ratio with the sample volume. The loading buffer was supplemented with 20 mM ethylenediaminetetraacetic acid (EDTA) as well as minor amounts of bromophenol blue (bpb) and xylene cyanol (xc) as color markers. In case of fluorescent samples (*e.g.* click reactions) unstained loading buffer was employed. Samples were heated to 95 °C for 2 min prior to gel loading.

Native PAGE was performed using a bpb/xc-stained 50 % glycerol solution as loading buffer without preheating of the samples.

DPAGE analyses were run at 300 V for 30 min up to 1 h, preparative gels were run at 400 V for 4 h, both at ambient temperature. Native PAGE was performed at 60 V for 3 h, 4 °C.

Analytical PA gels were stained with *SYBR® Safe* and visualized by UV illumination using a *Gel Doc 2000*, a *Genoplex*, or a *BioDoc-It™* gel documentation system. Fluorescently labeled nucleic acids were scanned using *Phosphorimager FLA-3000* with an excitation wavelength of λ_{ex} = 473 nm and an emission wavelength of λ_{em} = 520 nm beforehand. Precise composition of PA gels can be found in Table 23. Components **A**, **B**, and **C** were mixed as indicated; polymerization was initiated with ammonium peroxodisulfate (APS) and tetramethylethylenediamine (TEMED). Gel solutions were casted and allowed to polymerize for at least 30 min at room temperature.

Table 23. Composition of PA gels (5 mL total volume for analytical gels, 50 mL for preparative gels); DPAGE: **1** = 25 % acrylamide/bisacrylamide 19:1, 50 % urea, **2** = 8.3 M urea, **3** = 8.3 M urea in 10×TBE; native PAGE: **1** = 30 % acrylamide/bisacrylamide 37.5:1, **2** = ddH_2O, **3** = 10×TB.

Gel type	**1** [mL]	**2** [mL]	**3** [mL]	10 % APS [µL]	TEMED [µL]
8 % DPAGE, analytical	1.6	2.9	0.5	40	2
10 % DPAGE, analytical	2	2.5	0.5	40	2
12 % DPAGE, analytical	2.4	2.1	0.5	40	2
20 % DPAGE, analytical	4	0.5	0.5	40	2
20 % native PAGE, analytical	3.33	1.17	0.5	40	2
12 % DPAGE, preparative	24	21	5	400	20
20 % DPAGE, preparative	40	5	5	400	20

Agarose gels

PCR products were analyzed via agarose gel electrophoresis. 2 % agarose solutions (*w*/*v*) in 0.5×TBE were heated to near-boiling temperatures in a microwave (900 V, 2 min) and let cool to ~65 °C. From 50 mL agarose solution supplemented with 5 µL of a 1 % ethidium bromide (EtBr) solution a 20 pocket gel was casted and solidified for a minimum of 30 min at room temperature. Samples (typically 5 µL) were mixed with an appropriate of 6×DNA loading buffer. Gels were run for 13 min at 150 V before analyzing by UV exposure in a gel documentary system.

5.3.11 Purification of DNA and RNA

PCR products and *in vitro* transcribed RNA oligonucleotides were purified applying a variety of techniques, depending on the downstream implementation.

G-25 spin columns

For quick removal of salts (*e.g.* for HPLC-MS analysis), excess triphosphates (*e.g.* prior nucleoside digestions) or excess reagents (*e.g.* click reactions), *G-25 Illustra Microspin* columns (*GE Healthcare*) were employed according manufacturers' recommendations; a maximum of 40 µL crude transcription reactions were applied per column.

Silica-membrane spin columns

Where stated, PCR products were purified with *NucleoSpin Gel and PCR Clean-up kit* (*Macherey-Nagel*) according to the manufacturers' protocol and eluted into water (2×25 µL, 5 min incubation at 70 °C prior centrifugation) in order to determine nucleic acid concentration and/or remove primers and salts.

Centrifugal filters

For quick salt removal of RNA samples, *Amicon 3K* centrifugal filters (*Merck*) were employed. Oligonucleotide solutions were applied and centrifuged at 14000×g for 30 min. ddH_2O (500 µL) was added and passed through the membrane by centrifugation at 14000×g for 30 min. This step was repeated two to three times. Elution was carried out by upside-down centrifugation at 1000×g for 2 min. A second elution step was performed after careful rinsing of the membranes with ddH_2O (50 µL).

PAGE/Electroelution/EtOH precipitation

RNA transcripts were frequently purified via a gel elecrophoresis, elution, and precipitation cascade to remove *e.g.* truncated transcription products (for example templates **DNA^{NaM}**, **DNA^{sc_NaM}**, **DNA^{sc_ext_NaM}**). For this, samples were subjected to preparative PAGE (see section 5.3.10 for parameters), bands of interest were visualized by UV-shadowing on TLC plates, excised from the gel and eluted into an NH_4OAc solution (8 M, 170 µL) using an internally manufactured electroelution chamber (see Figure 155) filled with 1× TBE buffer at 150 V for 45 min–1 h. RNAs were precipitated by adding 1.2-1.7 mL ice-cold $EtOH_{abs.}$, thorough mixing, and overnight storage at -20 °C. Pelleting of the RNA by centrifugation (14000×g, 4 °C) was followed by removal of the supernatant. Optional washing with 200 µL 70 % EtOH, freeze drying and redissolving the samples in an appropriate volume of ddH_2O (e.g. 20 µL for RNA originating from a 100 µL transcription reaction) afforded highly pure nucleic acids.

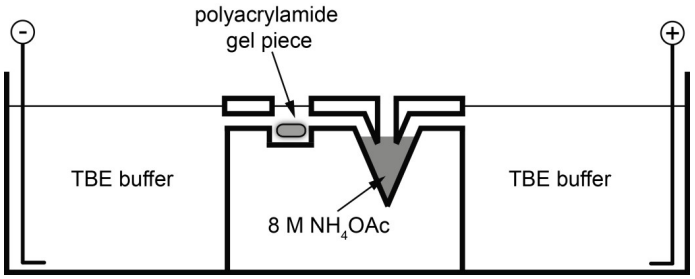


Figure 155. Electroelution chamber, illustration adapted from C. Mulhardt[429].

Crush and soak purification of DPAGE-separated oligonucleotides

For mild elution, bands of interest were excised from prep. DPAGE separation, crushed with a sterile pipette tip, and covered with an appropriate 1×buffer (tris or HEPES). Samples were agitated overnight at 8 °C and transferred into a sterile 5 mL syringe prepared with a plug of glass wool. Desiccation by pressure was followed by washings of the residue (2×1 mL ddH_2O). The combined eluates were freeze-dried and redissolved in an adequate volume of water (half the volume of initial transcription scale or 1 mL ddH_2O if desalination was to be applied).

Desalination using *Sep Pak* cartridges

If necessary, removal of salts from RNA transcripts was achieved with C18 *Sep Pak* cartidges (*Waters).* All equilibrating and washing steps were performed using sterile syringes. The cartridges were initially washed with HPLC grade MeCN (10 mL) and ddH_2O (20 mL). Samples, usually dissolved in 0.9-1.0 mL volume, were loaded and the flowthrough was reapplied four times to ensure maximum binding of target sequences. Washing with ddH_2O (20 mL) was followed by elution with 70 % MeCN (2 mL) and freeze-dried before redissolving in an appropriate volume of ddH_2O.

HPLC purification of *in vitro* transcribed RNA

To obtain analytically pure RNA transcripts (*e.g.* for EPR spectroscopy) micro purification via HPLC was employed. Typically, a 100 µL reaction was passed through a sterile 0.22 µm syringe filter. The filter was rinsed with ddH_2O (2×100 µL) and the filtrate concentrated to ≤100 µL by lyophilization before subjecting to HPLC on an *Agilent 1100* or an *Agilent 1260 Infinity II* Series HPLC system with an *EC 150/ 4.6 Nucleodur 100-5 C_{18} ec* column. Gradients used were 0→15 % B in 15 min with 0.1 M triethylammonium acetate (TEAAc) as solvent A, 0 % B for 5 min followed by a 0→30 % B gradient in 30 min with solvent A being 50 mM TEAAc, or 0→15 % B in 30 min with 0.1 % NH_4OAc (*w*/*v*) as solvent A, all at a flow rate of 1 mL min^{-1}. Acetonitrile was employed as solvent B in all cases. The fractions of interest were collected in 1.5 mL or 15 mL tubes, freeze dried, and redissolved in appropriate volumes of ddH_2O.

5.3.12 Concentration determination of nucleic acids

To determine the concentration of purified DNA or RNA solutions the absorption at 260 nm (A_{260}) was measured using a *Nanodrop 2000c* UV-spectrometer. Concentrations were obtained from the A_{260} value and software-assisted calculation with *Oligonucleotide Properties Calculator.*[384] Native sequences containing canonical bases were plotted for unnatural base-modified oligonucleotides.

Table 24. Single values of DPAGE/electroelution-purified RNA transcripts (see 3.1.2.4).

	RNA^{n}	**RNA^{NaM}**
	(per 100 µL transcription)	
	4.89	1.04
n	4.89	1.04
[nmol]	2.37	0.69
	-	0.75
Ø	4.05	0.88
SD	1.19	0.16
norm.	**100 %**	**22 %**

Table 25. Single values comparing incorporation efficiency of r**TPT3^{CP}** TP (**34**) and r**TPT3^{NOR}** TP (**29**) during *in vitro* transcription of **DNA^{NaM}** (see 3.1.2.4); A_{260} values are averaged from four measurements. Amount of RNA obtained (*n* [nmol]) correspond to 36 µL reaction volume, purified via DPAGE/electoelution.

	r**TPT3^{CP}** TP (**34**)		r**TPT3^{NOR}** TP (**29**)	
	30 min	5 h	30 min	5 h
	2.01	5.85	1.08	7.78
A_{260}	1.24	9.09	0.64	2.90
	2.34	5.72	1.13	3.62
Ø	1.87 ± 0.56	6.88 ± 1.91	0.95 ± 0.27	4.76 ± 2.63
***c* [µM]**	**8.72** ± 2.61	**32.08** ± 8.90	**4.43** ± 1.26	**22.19** ± 12.26
***n* [nmol]**	**0.17** ± 0.054	**0.64** ± 0.18	**0.09** ± 0.03	**0.44** ± 0.25
norm.	**100 %**	**100 %**	**53 %**	**69 %**

5.3.13 HPLC-MS analysis of DNA and RNA

HPLC-ESI mass spectrometry of RNA transcripts was performed on an *HTC esquire* from *Bruker Daltonic* in combination with an *Agilent 1100 Series* HPLC system (*Agilent Technologies*) using 10 mM NEt_3/100 mM HFIP as solvent A and MeCN as solvent B. Gradients employed were either 3→20 % B in 20 min at a flow rate of 0.4 mL min^{-1} on a *Zorbax StableBond* (80 Å C18, 2.1×50 mm, 5 µm) column (*Agilent Technologies*) or 3 % B for 9 min followed by a from 3→40 % B gradient in 11 min at a flow rate of 0.2 mL min^{-1} on a *XTerra® MS* (2.1×100 mm, 5 µm) C_{18} column (*Waters*), both elution parameters being of sufficient separation quality.

Crude, DNase-digested transcription reactions (100 µL scale) were generally desalted by gel filtration using *G-25* columns or (if not purified by HPLC). Samples of approx. 20 pmol were subjected to HPLC-MS analysis.

Crude ESI$^-$ spectra obtained by integration from suitable time frames within the respective UV trace were evaluated via *Data Analysis* using default deconvolution parameters (1000–40000 mass units, max. charge = 60, max. number of compounds = 4, min. 3 peaks per compound) or a suitable neutral mass area of usually 5000–15000 mass units, max. charge = 60, max. number of compounds = 10, min. 5 peaks per compound.
For the analysis of of *glmS* ribozyme variants (section 3.1.2.2) parameters were as follows: max. charge = 80, max. number of compounds = 15, min. 4–6 peaks per compound, and depending on the respective fragment an ample neutral mass area (5000–15,000 m/z for cleaved fragments, 47,500–52,500 neutral mass units for cleaved ribozymes, and 57,500–62,500 mass units for full ribozymes).

Table 26. Incorporation efficiency of r**TPT3^{CP}** TP (**34**) via T7 *in vitro* transcription from **DNA^{NaM}**.

	Area full-length transcript **RNA^{CP}** [%]	Area truncated transcript [%]
Single values	91.34	8.66
	84.44	15.56
	89.95	10.05
	87.66	12.34
	83.27	16.73
Mean	87.33	12.67
SD	3.09	3.09

5.3.14 Ribozyme cleavage assays

For cleavage assays 14 µL crude, DNase-digested *glmS* RNA **glmS^{CP}** with **TPT3^{CP}** modifications was incubated with 3.5 µL **^{H}Tet-BDP** (**57**, 500 µM in DMSO) for 1 h at room temperature under exclusion of light.
The BODIPY-labeled *glmS* ribozyme **glmS^{CP}-Tet-BDP** (in total 17.5 µL) was heated to 95 °C for 2 min in 1× cleavage buffer (see

Table 8) and incubated at room temperature for 5 min. Glucosamine-6-phosphate (GlcN6P, **1**) was then added to give a final concentration of 5 µM to start the cleavage reaction. Samples of 5 µL (corresponding to 2 µL transcript) were taken at 1, 2, 5, 15, 30 and 60 min, and mixed thoroughly with 5 µL loading buffer inhibiting further ribozyme cleavage. Resulting RNAs were analyzed via 12 % DPAGE and quantified by fluorescence scanning on a phosphorimager (λ_{ex} = 475 nm, λ_{em} = 520 nm). The areas of interest were integrated with *AIDA* software and the data obtained were evaluated with *Origin8*. The reaction progress of

the cleavage reaction was fitted to a single exponential function ($y = y_0 + A(1 - exp^{-rx}$). Table 27 shows all background corrected band intensity values and resulting cleaved fractions of *glmS* assays.

Table 27. Raw intensity data of fluorescent band areas obtained from **glmS^{CP}-Tet-BDP** cleavage assays, analyzed via 12 % DPAGE (background subtracted).

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	199.3	186.0	48.3	90.4	76.4	45.8
1	176.6	279.7	61.3	134.2	133.4	49.9
2	149.0	311.3	67.6	104.2	137.2	56.9
5	114.3	314.3	73.3	87.4	130.5	59.9
15	83.8	312.6	78.9	79.9	131.4	62.2
30	77.7	302.0	79.5	77.0	137.1	64.0
60	79.1	318.8	80.1	87.6	144.0	62.2

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	105.0	86.7	45.2	2016.3	1215.3	37.6
1	160.5	172.6	51.8	1541.1	1356.7	46.8
2	142.6	177.5	55.5	1525.4	1648.6	51.9
5	116.6	159.9	57.8	1248.1	1820.7	59.3
15	117.5	172.3	59.5	983.0	2179.0	68.9
30	101.7	161.7	61.4	975.8	2187.9	69.2
60	86.2	149.4	63.4	919.0	2152.9	70.1

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	2247.5	1377.0	38.0	1148.3	748.7	39.5
1	1800.4	1368.8	43.2	1106.0	1016.2	47.9
2	1510.2	1593.8	51.3	887.1	968.7	52.2
5	1360.2	2002.9	59.6	756.2	1229.6	61.9
15	1077.4	2231.2	67.4	600.2	1129.0	65.3
30	970.6	2075.6	68.1	544.9	1116.2	67.2
60	956.3	2168.1	69.4	576.5	1216.2	67.8

Table 27 (continued)

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	836.3	637.2	43.2	1344.8	934.3	41.0
1	1142.9	979.0	46.1	1237.9	1069.0	46.3
2	944.1	1055.8	52.8	1093.4	1187.1	52.1
5	718.5	1180.1	62.2	853.0	1382.4	61.8
15	624.4	1294.8	67.5	649.3	1516.0	70.0
30	580.6	1182.1	67.1	575.2	1367.8	70.4
60	513.5	1023.1	66.6	542.6	1398.3	72.0

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	1026.2	684.6	40.0	1671.4	1184.0	41.5
1	1041.2	1045.4	50.1	1413.2	1263.5	47.2
2	881.2	1123.4	56.0	1223.2	1465.0	54.5
5	730.5	1239.9	62.9	1040.4	1720.8	62.3
15	520.8	1133.8	68.5	945.8	2107.9	69.0
30	537.3	1140.4	68.0	724.2	1837.6	71.7
60	503.5	1109.2	68.8	914.3	1968.1	68.3

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	1289.1	1114.9	46.4	1760.8	1214.8	40.8
1	1706.2	1455.9	46.0	1321.6	1412.9	51.7
2	1432.2	1711.3	54.4	1165.1	1779.5	60.4
5	1080.8	1872.3	63.4	893.1	1791.4	66.7
15	829.1	1857.4	69.1	751.0	1844.2	71.1
30	883.5	2016.0	69.5	770.0	1840.7	70.5
60	760.5	2000.5	72.5	634.4	1672.0	72.5

t [min]	Int. full ribozyme [a.u.]	Int. cleaved ribozyme [a.u.]	fraction cleaved [%]
0	1307.0	1041.9	44.4
1	1163.9	1128.5	49.2
2	964.9	1250.9	56.5
5	721.1	1377.8	65.6
15	663.9	1489.8	69.2
30	615.5	1456.3	70.3

5.3.15 Assessing unspecific **TPT3^{CP}** nucleotide incorporation

Transcription reactions were carried out in 20 µL scale (triplicates) with either d**NaM**-modified (**DNA^{NaM}**) or unmodified template **DNA^{n}** with varying concentrations of r**TPT3^{CP}** TP (**34**, 0 mM and 1.0 mM for **DNA^{NaM}**, 0.0, 0.1, 0.5, and 1.0 mM for native **DNA^{n}**. Crude transcripts were DNase I digested and purified via *G-25* columns to remove excess **34**. The concentration was determined; samples of 150 pmol were prepared in 30 µL ddH_2O and subsequently reacted with 10 eq. of **^{H}Tet-BDP** (**57**). Excess **57** was removed using *G-25* columns. For each sample 25 µL eluate were transferred to a 96-well plate and fluorescence intensity was measured in a plate reader (*Tecan Ultra*, λ_{ex} = 485 nm, λ_{em} = 535 nm). Crude data were corrected by the background signal and normalized to the positive control (**RNA^{CP}**, 1 mM **34**, set as relative fluorescence = 100).

Table 28. Fluorescence values assessing unspecific r**TPT3^{CP}** nucleotide incorporation; data are the average of at least two independent measurements, background-subtracted and normalized.

	RNA^{n1}		**RNA^{C}**			**RNA^{CP}**	
c [mM]	0.0	0.1	0.5	1.0	2.0	0.0	1.0
Fl. Int. [a.u.]	1.669	1.662	1.617	1.644	1.782	1.121	100.000
SD	0.158	0.090	0.183	0.123	0.042	0.475	26.267

5.3.16 CD spectroscopy

Oligonucleotides (5 µM) were hybridized in 1× phosphate buffer by heating to 70 °C for 5 min, followed by a gradient 70→18 °C, cooling rate 2 °C min^{-1}. CD experiments were recorded on a *JASCO J-810* spectropolarimeter at 25 °C (0.1 cm high precision cell) by averaging 10 scans (220–340 nm, 100 nm min^{-1}, response time = 0.1 s). Buffer spectra were subtracted from the data obtained.

5.3.17 Thermal denaturation experiments

Thermal denaturation experiments of modified and unmodified oligonucleotides were carried out on a *Cary 100 UV-Vis* spectrophotometer. 1 µM samples were prepared in 100 µL 1× phosphate buffer, occasionally with 10 mM $MgCl_2$, annealed 5 min at 70 °C followed by cooling to 18 °C with 2 °C min^{-1}, and analyzed in micro-cuvettes at 265 nm. The temperature range for melting curve measurements was set from 12 °C to 85 °C or 20 °C to 95 °C with a rate of 1.0 °C min^{-1}. Melting points were determined from three independent measurements.

Table 29. Melting points of RNA duplex sequences presented in this study, each value is the average of at least three independent measurements, errors are SDs thereof; *with 10 mM $MgCl_2$, temperature range 20–95 °C; ** without $MgCl_2$, temperature range 12–85 °C.

RNA	T_m [°C]
RNA^{sc_n}	67.1±0.3*
RNA^{sc_CP}	68.4±0.5*
RNA^{sc_NO}	56.0±0.2**
RNA^{sc_n}	75.8±0.8**
RNA^{sc_ext_NO}	65.2±0.4**
RNA^{sc_ext_n}	77.6±1.3**

5.3.18 EPR sample preparation

Sample preparation for *cw*-X-band EPR measurements

RNA oligonucleotides were dissolved in 10 µL 1× phosphate buffer and hybridized by heating to 70 °C for 5 min, followed by a gradient 70→18 °C with a cooling rate of 2 °C min^{-1}. Transferred into 10 µL glass capillaries, *cw*-X-band spectra were recorded.

Sample preparation for PELDOR measurements

RNA oligonucleotides were dissolved in 50 µL or 80 µL 1× phosphate buffer and hybridized by heating to 70 °C for 5 min, followed by a gradient 70→18 °C with a cooling rate of 2 °C min^{-1}. For PELDOR measurements, RNA solutions were lyophilized after hybridization and then dissolved in sterile filtered D_2O supplemented with 20 % deuterated ethylene glycol. The samples were transferred to a 3 mm quartz Q-band EPR tube and flash-cooled in liquid nitrogen prior PELDOR measurements. The RNA amounts employed for EPR experiments are stated within the corresponding paragraph in section 3.2.2.

5.3.19 List of DNA and RNA sequences

The following section provides all unmodified and modified synthetic DNA templates (Table 30), full-length dsDNA templates generated by assembly or plasmid PCR (Table 31), chemically synthesized RNA oligonucleotides as well as and sequences resulting from *in vitro* transcription (Table 32).

Table 30. List of chemically synthesized DNA templates used in this study, underlined segments of *glmS* templates highlight overlapping or primer binding regions.

Oligonucleotide	Sequence
DNA^{n1}	5′-CAC TAC TCG GGA TTC CCT ATA GTG AGT CGT ATT AT-3′
DNA^{NaM}	5′-CAC Td**NaM**C TCG GGA TTC CCT ATA GTG AGT CGT ATT AT-3′
tDNA^{n}	5′-TGG TCC GGC GGG CCG GAT TTG AAC CAG CGC CAT GCG GAT TAA GAG TCC GCC GTT CTG CCC TGC TGA ACT ACC GCC GGT ATA GTG AGT CGT ATT ATC-3′
tDNA^{NaM}	5′-TGG TCC GGC GGG CCG GAT TTG AAC CAG CGC CAT GCG GAT Td**NaM**A GAG TCC GCC GTT CTG CCC TGC TGA ACT ACC GCC GGT ATA GTG AGT CGT ATT ATC-3′
tDNA^{RV}	5′-GAT AAT ACG ACT CAC TAT ACC GGC GGT AGT TCA GCA GGG CAG AAC GGC GGA CTC TAA ATC CGC ATG GCG CTG GTT CAA ATC CGG CCC GCC GGA CCA-3′
Primer^{n1}	5′-ATA ATA CGA CTC ACT ATA GGG-3′
glmS_Primer_FW^{1}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT A-3′
glmS_Primer_RV^{1}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GA-3′
DNA_glmS^{1n}	5′-CCT CCA TCC TCG TCA ACT AAG CCT TTT TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′
DNA_glmS^{1NaM}	5′-CCT CCA TCC TCG TCA ACT AAG CCT Td**NaM**T TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′
DNA_glmS^{2n}	5′-GCT TAG TTG ACG AGG ATG GAG GTT ATC GAA TTT TCG GCG GAT GCC TCC CGG CTG AGT GTG CAG ATC ACA GCC GTA AGG ATT TCT TC-3′
DNA_glmS^{3n}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
DNA_glmS^{3NaM}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCd**NaM** CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
Xist_Primer_FW	5′-TAA TAC GAC TCA CTA TAG GTC CCC GCC ATT CCA TGC CCA ACG GG-3′
Xist_Primer_RV^{n}	5′-GGT AGT GCA ATA ACT CAC AAA ACC ATA TTT CCA TCC ACC AAG CG-3′
Xist_Primer_RV^{NaM}	5′-GGT AGT GCA ATA ACT CAC AAA ACC ATA Td**NaM**T CCA TCC ACC AAG CG-3′

Table 30 (continued).

Primer^{n2}	5′-ATA ATA CGA CTC ACT ATA GG-3′
DNA^{sc_n}	5′-GGA TCT GAT ATC AGA TCC TA TAG TGA GTC GTA TTA T-3′
DNA^{sc_NaM}	5′-GGd**NaM** TCT GAT ATC AGA TCC TA TAG TGA GTC GTA TTA T-3′
DNA^{sc_ext_NaM}	5′-GGd**NaM** TCT GAT GCA TCA GAT CCT ATA GTG AGT CGT ATT AT-3′
DNA^{sc_NaM_OMe}	5′-(OMe)G(OMe)Gd**NaM** TCT GAT ATC AGA TCC TA TAG TGA GTC GTA TTA T-3′
glmS_Primer_FW^{short}	5′-TAA TAC GAC TCA CTA TAG GTA TAA TTA TAG-3′
glmS_DNA^{n_short_RV}	5′-GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT ACC TAT AGT GAG TCG TAT TA-3′
glmS_DNA^{NaM_P1P4a_short_RV}	5′-GAT TTC TCT TTG TTC Ad**NaM**G GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAd**NaM** GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT ACC TAT AGT GAG TCG TAT TA-3′
glmS_DNA^{NaM_P1P4L_short_RV}	5′-GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA Gd**NaM**A ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAd**NaM** GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT ACC TAT AGT GAG TCG TAT TA-3′
glmS_Primer_FW^{2}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC-3′
glmS_Primer_RV^{2}	5′-AGA TCA TGT GAT TTC TCT TTG TTC-3′
glmS_Primer_RV^{NaM_4R}	5′-AGA TCA TGT GAT Td**NaM**C TCT TTG TTC-3′
glmS_DNA^{1NaM_P1}	5′-CCT CCA TCC TCG TCA ACT Ad**NaM**G CCT TTT TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′
glmS_DNA^{1n_2}	5′-CCT CCA TCC TCG TCA ACT AAG CCT TTT TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′
glmS_DNA^{2n_2}	5′-AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT C-3′

Table 30 (continued).

glmS_DNA^{3NaM_P4a}	5′-AGA TCA TGT GAT Td**NaM**C TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
glmS_DNA^{3NaM_P4R}	5′-AGA TCA TGT GAT TTC TCT TTG TTC Ad**NaM**G GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
glmS_DNA^{3NaM_P4RP4a}	5′-AGA TCA TGT GAT Td**NaM**C TCT TTG TTC Ad**NaM**G GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
glmS_DNA^{3n_2}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG-3′
glmS_DNA^{3n_P4RP4a_HH}	5′-CAT CCG GTG ACT CAT GTG ATT TCT CTT TGT TCA AGG AGT CAC CCC CTT GGT TTG AAG AAA TCC TTA CGG CTG TG-3′
glmS_DNA^{3NaM_P4RP4a_HH}	5′-CAT CCG GTG ACT CAT GTG ATT Td**NaM**T CTT TGT TCA d**NaM**GG AGT CAC CCC CTT GGT TTG AAG AAA TCC TTA CGG CTG TG-3′
glmS_DNA^{4n_HH}	5′-ATC ACA TGA GTC ACC GGA TGT GCT TTC CGG TCT GAT GAG TCC GTG AGG ACG AAA CTC AT-3′
glmS_Primer_RV^{HH}	5′-ATG AGT TTC GTC CTC ACG GA-3′
T7 Primer	5′-GGC GTA ATA CGA CTC ACT ATA GGG-3′
DNA^{n_RT}	5′-CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC CAT CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CCT ATA GTG AGT CGT ATT ACG CC-3′
DNA^{5SICS_RT}	5′-CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC Cd**5SICS**T CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CCT ATA GTG AGT CGT ATT ACG CC-3′
DNA^{NaM_RT}	5′-CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC Cd**NaM**T CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CCT ATA GTG AGT CGT ATT ACG CC-3′
Primer^{RT}	5′-(6-FAM)CGA AGC TTA CTT GTC GTC GTC GTC CTT GT-3′

Table 31. List of DNA (template) sequences prepared by PCR or rtc.

Oligonucleotide	Sequence
glmS^{n}_DNA^{full_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCT TGA ACA AAG AGA AAT CAC ATG ATC T-3′
glmS^{n}_DNA^{full_anti}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′
glmS_DNA^{full_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA Ad**TPT3**A AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG Gd**TPT3**G ACT CCT TGA ACA AAG AGA AAT CAC ATG ATC T-3′
glmS_DNA^{full_anti}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCd**NaM** CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTd**NaM** TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′

Table 31 (continued).

Xist^{n}_DNA^{anti}	5′-GGT AGT GCA ATA ACT CAC AAA ACC ATA TAT CCA TCC ACC AAG CGC CCC GTT GGG CCG TGA AAA AAA AAA TTT AAA GCA GGT ATC CAC AGC CCC GAT GGG CAA AAG AAA AAG AAA AAA AAA TAA TAA CAG CAG GTA TCC GAG GCC CCG TTG GGC ATG GGA AAA AAA GAC TAA ACG CAG GTA TCC GAG GTC CCG ATG GAC CGA GAA AGG TTT TTT TGT TTT TTT TTT TTT TTT TAC AAA AAG CAG GTA TCC ATG GCC CCG ATG GGC TAA GGA GAA GAA AAA AAG AAT AAA AGC AGG TAT CCA CAG CCC AGA TGG GCA AGT TTA GAA AAA AAA ATA ATA AGA AAA AAA AAG AAT GAA AAG GCA GGT AAG TAT CCA AAA CCC CGT TGG GCA TGG AAT GGC GGG GAC CTA TAG TGA GTC GTA TTA-3′
Xist^{n}_DNA^{sense}	5′-TAA TAC GAC TCA CTA TAG GTC CCC GCC ATT CCA TGC CCA ACG GGG TTT TGG ATA CTT ACC TGC CTT TTC ATT CTT TTT TTT TCT TAT TAT TTT TTT TTC TAA ACT TGC CCA TCT GGG CTG TGG ATA CCT GCT TTT ATT CTT TTT TTC TTC TCC TTA GCC CAT CGG GGC CAT GGA TAC CTG CTT TTT GTA AAA AAA AAA AAA AAA ACA AAA AAA CCT TTC TCG GTC CAT CGG GAC CTC GGA TAC CTG CGT TTA GTC TTT TTT TCC CAT GCC CAA CGG GGC CTC GGA TAC CTG CTG TTA TTA TTT TTT TTT CTT TTT CTT TTG CCC ATC GGG GCT GTG GAT ACC TGC TTT AAA TTT TTT TTT TCA CGG CCC AAC GGG GCG CTT GGT GGA TGG AAA TAT GGT TTT GTG AGT TAT TGC ACT ACC-3′
Xist^{CP}_DNA^{anti}	5′-GGT AGT GCA ATA ACT CAC AAA ACC ATA Td**NaM**T CCA TCC ACC AAG CGC CCC GTT GGG CCG TGA AAA AAA AAA TTT AAA GCA GGT ATC CAC AGC CCC GAT GGG CAA AAG AAA AAG AAA AAA AAA TAA TAA CAG CAG GTA TCC GAG GCC CCG TTG GGC ATG GGA AAA AAA GAC TAA ACG CAG GTA TCC GAG GTC CCG ATG GAC CGA GAA AGG TTT TTT TGT TTT TTT TTT TTT TTT TAC AAA AAG CAG GTA TCC ATG GCC CCG ATG GGC TAA GGA GAA GAA AAA AAG AAT AAA AGC AGG TAT CCA CAG CCC AGA TGG GCA AGT TTA GAA AAA AAA ATA ATA AGA AAA AAA AAG AAT GAA AAG GCA GGT AAG TAT CCA AAA CCC CGT TGG GCA TGG AAT GGC GGG GAC CTA TAG TGA GTC GTA TTA-3′

Table 31 (continued).

Xist^{CP}_DNA^{sense}	5′-TAA TAC GAC TCA CTA TAG GTC CCC GCC ATT CCA TGC CCA ACG GGG TTT TGG ATA CTT ACC TGC CTT TTC ATT CTT TTT TTT TCT TAT TAT TTT TTT TTC TAA ACT TGC CCA TCT GGG CTG TGG ATA CCT GCT TTT ATT CTT TTT TTC TTC TCC TTA GCC CAT CGG GGC CAT GGA TAC CTG CTT TTT GTA AAA AAA AAA AAA AAA ACA AAA AAA CCT TTC TCG GTC CAT CGG GAC CTC GGA TAC CTG CGT TTA GTC TTT TTT TCC CAT GCC CAA CGG GGC CTC GGA TAC CTG CTG TTA TTA TTT TTT TTT CTT TTT CTT TTG CCC ATC GGG GCT GTG GAT ACC TGC TTT AAA TTT TTT TTT TCA CGG CCC AAC GGG GCG CTT GGT GGA TGG AAA Td**TPT3**T GGT TTT GTG AGT TAT TGC ACT ACC-3′
glmS^{n2}_DNA^{full_anti}	5′-AGA TCA TGT GAT TTC TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′
glmS^{n2}_DNA^{full_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCT TGA ACA AAG AGA AAT CAC ATG ATC T-3′
glmS_DNA^{P1P4afull_anti}	5′-AGA TCA TGT GAT TTC TCT TTG TTC Ad**NaM**G GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAd**NaM** GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′
glmS_DNA^{P1P4afull_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG Cd**TPT3**T AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCd**TPT3** TGA ACA AAG AGA AAT CAC ATG ATC T-3′

Table 31 (continued).

glmS_DNA^{P4RP4afull_anti}	5′-AGA TCA TGT GAT Td**NaM**C TCT TTG TTC Ad**NaM**G GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′
glmS_DNA^{P4P4afull_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCd**TPT3** TGA ACA AAG AGd**TPT3** AAT CAC ATG ATC T-3′
glmS_DNA^{P4Rfull_anti}	5′-AGA TCA TGT GAT Td**NaM**C TCT TTG TTC AAG GAG TCA CCC CCT TGG TTT GAA GAA ATC CTT ACG GCT GTG ATC TGC ACA CTC AGC CGG GAG GCA TCC GCC GAA AAT TCG ATA ACC TCC ATC CTC GTC AAC TAA GCC TTT TTC CGG GCG GCT TAG TTC GGG CGC TAT AAT TAT AGG TAA AGC AAT AAT CCT ATA GTG AGT CGT ATT A-3′
glmS_DNA^{P4Rfull_sense}	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCT TGA ACA AAG AGd**TPT3** AAT CAC ATG ATC T-3′
glmS_DNA^{n}–HH_full_anti	5′-ATG AGT TTC GTC CTC ACG GAC TCA TCA GAC CGG AAA GCA CAT CCG GTG ACT CAT GTG ATT **T**CT CTT TGT TCA **A**GG AGT CAC CCC CTT GGT TTG AAG AAA TCC TTA CGG CTG TGA TCT GCA CAC TCA GCC GGG AGG CAT CCG CCG AAA ATT CGA TAA CCT CCA TCC TCG TCA ACT AAG CCT TTT TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′

Table 31 (continued).

glmS_DNA^{n}_HH_full_sense	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CC**T** TGA ACA AAG AG**A** AAT CAC ATG AGT CAC CGG ATG TGC TTT CCG GTC TGA TGA GTC CGT GAG GAC GAA ACT CAT-3′
glmS_DNA^{P4RP4a}_HH_full_anti	5′-ATG AGT TTC GTC CTC ACG GAC TCA TCA GAC CGG AAA GCA CAT CCG GTG ACT CAT GTG ATT d**NaM**CT CTT TGT TCA d**NaM**GG AGT CAC CCC CTT GGT TTG AAG AAA TCC TTA CGG CTG TGA TCT GCA CAC TCA GCC GGG AGG CAT CCG CCG AAA ATT CGA TAA CCT CCA TCC TCG TCA ACT AAG CCT TTT TCC GGG CGG CTT AGT TCG GGC GCT ATA ATT ATA GGT AAA GCA ATA ATC CTA TAG TGA GTC GTA TTA-3′
glmS_DNA^{P4RP4a}_HH_full_sense	5′-TAA TAC GAC TCA CTA TAG GAT TAT TGC TTT ACC TAT AAT TAT AGC GCC CGA ACT AAG CCG CCC GGA AAA AGG CTT AGT TGA CGA GGA TGG AGG TTA TCG AAT TTT CGG CGG ATG CCT CCC GGC TGA GTG TGC AGA TCA CAG CCG TAA GGA TTT CTT CAA ACC AAG GGG GTG ACT CCd**TPT3** TGA ACA AAG AGd**TPT3** AAT CAC ATG AGT CAC CGG ATG TGC TTT CCG GTC TGA TGA GTC CGT GAG GAC GAA ACT CAT-3′
cDNA^{n}_RT	5′-(6-FAM)CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC CAT CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CC-3′ DNA
cDNA^{NaM}_RT	5′-(6-FAM)CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC Cd**NaM**T CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CC-3′ DNA
cDNA^{TPT3}_RT	5′-(6-FAM)CGA AGC TTA CTT GTC GTC GTC GTC CTT GTA GTC Cd**TPT3**T CTT GTA CTT TTT GTA CAT GTT TTT CTC CTT CTT AAA GTT AAC CC-3′ DNA

Table 32. List of RNA sequences used in this study.

Oligonucleotide	Sequence
RNA^{n1}	5′-GGG AAU CCC GAG UAG UG-3′
RNA^{CP}	5′-GGG AAU CCC GAG r**TPT3^{CP}**AG UG-3′
tRNA^{n}	5′-CCG GCG GUA GUU CAG CAG GGC AGA ACG GCG GAC UCU UAA UCC GCA UGG CGC UGG UUC AAA UCC GGC CCG CCG GAC CA-3′
tRNA^{CP}	5′-CCG GCG GUA GUU CAG CAG GGC AGA ACG GCG GAC UCU r**TPT3^{CP}**AA UCC GCA UGG CGC UGG UUC AAA UCC GGC CCG CCG GAC CA-3′
glmS_RNA^{n1}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC UUG AAC AAA GAG AAA UCA CAU GAU CU-3′
glmS_RNA^{CP}1	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGr**TPT3^{CP}** GAC UCC UUG AAC AAA GAG AAA UCA CAU GAU CU-3′
glmS_RNA^{CP}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAr**TPT3^{CP}** AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGr**TPT3^{CP}** GAC UCC UUG AAC AAA GAG AAA UCA CAU GAU CU-3′

Table 32 (continued).

Xist_RNA^{n}	5′-GGU CCC CGC CAU UCC AUG CCC AAC GGG GUU UUG GAU ACU UAC CUG CCU UUU CAU UCU UUU UUU UUC UUA UUA UUU UUU UUU CUA AAC UUG CCC AUC UGG GCU GUG GAU ACC UGC UUU UAU UCU UUU UUU CUU CUC CUU AGC CCA UCG GGG CCA UGG AUA CCU GCU UUU UGU AAA AAA AAA AAA AAA AAC AAA AAA ACC UUU CUC GGU CCA UCG GGA CCU CGG AUA CCU GCG UUU AGU CUU UUU UUC CCA UGC CCA ACG GGG CCU CGG AUA CCU GCU GUU AUU AUU UUU UUU UCU UUU UCU UUU GCC CAU CGG GGC UGU GGA UAC CUG CUU UAA AUU UUU UUU UUC ACG GCC CAA CGG GGC GCU UGG UGG AUG GAU AUA UGG UUU UGU GAG UUA UUG CAC UAC C-3′
Xist_RNA^{CP}	5′-GGU CCC CGC CAU UCC AUG CCC AAC GGG GUU UUG GAU ACU UAC CUG CCU UUU CAU UCU UUU UUU UUC UUA UUA UUU UUU UUU CUA AAC UUG CCC AUC UGG GCU GUG GAU ACC UGC UUU UAU UCU UUU UUU CUU CUC CUU AGC CCA UCG GGG CCA UGG AUA CCU GCU UUU UGU AAA AAA AAA AAA AAA AAC AAA AAA ACC UUU CUC GGU CCA UCG GGA CCU CGG AUA CCU GCG UUU AGU CUU UUU UUC CCA UGC CCA ACG GGG CCU CGG AUA CCU GCU GUU AUU AUU UUU UUU UCU UUU UCU UUU GCC CAU CGG GGC UGU GGA UAC CUG CUU UAA AUU UUU UUU UUC ACG GCC CAA CGG GGC GCU UGG UGG AUG GAr**TPT3^{CP}** AUA UGG UUU UGU GAG UUA UUG CAC UAC C-3′
RNA^{sc_n}	5′-GGA UCU GAU AUC AGA UCC-3′
RNA^{sc_CP}	5′-GGA UCU GAU AUC AGA r**TPT3^{CP}**CC-3′
RNA^{sc_NO}	5′-GGA UCU GAU AUC AGA r**TPT3^{NO}**CC-3′
RNA^{sc_ext_n}	5′-GGA UCU GAU GCA UCA GAU CC-3′
RNA^{sc_ext_NO}	5′-GGA UCU GAU GCA UCA GAr**TPT3^{NO}** CC-3′
RNA^{sc_NO} (truncated)	5′-GGA UCU GAU AUC AGA-3′
RNA^{sc_NO}-1	5′-GGA UCU GAU AUC AGA r**TPT3^{NO}**C-3′
RNA^{I}	5′-GGG AAU CCC GAG r**TPT3^{I}**AG UG-3′
RNA^{TPT3}	5′-GGG AAU CCC GAG r**TPT3**AG UG-3′
RNA^{NaM}	5′-GGG AAU CCC GAG r**NaM**AG UG-3′

Table 32 (continued).

glmS_RNA^{NO_n}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC UUG AAC AAA GAG AAA UCA CAU GAU CU-3′
glmS_RNA^{NO}	5′-G GAU UAU UGC UUU ACC UAU AAU UAU AGC GCC CGA ACU AAG CCG CCC GGA Ar**TPT3^{NO}**A AGG CUU AGU UGA CGA GGA UGG AGG UUA UCG AAU UUU CGG CGG AUG CCU CCC GGC UGA GUG UGC AGA UCA CAG CCG UAA GGA UUU CUU CAA ACC AAG GGG Gr**TPT3^{NO}**G ACU CCU UGA ACA AAG AGA AAU CAC AUG AUC U-3′
glmS_RNA^{n_short}	5′-GGU AUA AUU AUA GCG CCC GAA CUA AGC CGC CCG GAA AAA GGC UUA GUU GAC GAG GAU GGA GGU UAU CGA AUU UUC GGC GGA UGC CUC CCG GCU GAG UGU GCA GAU CAC AGC CGU UAG GAU UUC UUC AAA CCA AGG GGG UGA CUC CUU GAA CAA AGA GAA AUC-3′
glmS_RNA^{NO_P1P4a_short}	5′-GGU AUA AUU AUA GCG CCC GAA CUA AGC CGC CCG GAA AAA GGC r**TPT3^{NO}**UA GUU GAC GAG GAU GGA GGU UAU CGA AUU UUC GGC GGA UGC CUC CCG GCU GAG UGU GCA GAU CAC AGC CGU UAG GAU UUC UUC AAA CCA AGG GGG UGA CUC Cr**TPT3^{NO}**U GAA CAA AGA GAA AUC-3′
glmS_RNA^{NO_P1P4L_short}	5′-GGU AUA AUU AUA GCG CCC GAA CUA AGC CGC CCG GAA AAA GGC r**TPT3^{NO}**UA GUU GAC GAG GAU GGA GGU UAU CGA AUU UUC GGC GGA UGC CUC CCG GCU GAG UGU GCA GAU CAC AGC CGU UAG GAU Ur**TPT3^{NO}**C UUC AAA CCA AGG GGG UGA CUC CUU GAA CAA AGA GAA AUC-3′
glmS_RNA^{NO_P1P4a}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCr**TPT3^{NO}** UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC r**TPT3^{NO}**UG AAC AAA GAG AAA UCA CAU GAU CU-3′

Table 32 (continued).

glmS_RNA^{NO_P4RP4a}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC r**TPT3^{NO}**UG AAC AAA GAG r**TPT3^{NO}**AA UCA CAU GAU CU-3′
glmS_RNA^{NO_P4a}	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC r**TPT3^{NO}**UG AAC AAA GAG AAA UCA CAU GAU CU-3′
glmS_RNA$^{NO/CP_P4RP4a_HH}$	5′-GGA UUA UUG CUU UAC CUA UAA UUA UAG CGC CCG AAC UAA GCC GCC CGG AAA AAG GCU UAG UUG ACG AGG AUG GAG GUU AUC GAA UUU UCG GCG GAU GCC UCC CGG CUG AGU GUG CAG AUC ACA GCC GUA AGG AUU UCU UCA AAC CAA GGG GGU GAC UCC r**TPT3$^{NO/CP}$**UG AAC AAA GAG r**TPT3$^{NO/CP}$**AA UCA CAU GAG UCA CCG GAU GUG CUU UCC GGU CUG AUG AGU CCG UGA GGA CGA AAC UCA U-3′
RNA^{n_RT}	5′-GGG UUA ACU UUA AGA AGG AGA AAA ACA UGU ACA AAA AGU ACA AGA UGG ACU ACA AGG ACG ACG ACG ACA AGU AAG CUU CG-3′
RNA^{NaM_RT}	5′-GGG UUA ACU UUA AGA AGG AGA AAA ACA UGU ACA AAA AGU ACA AGA r**NaM**GG ACU ACA AGG ACG ACG ACG ACA AGU AAG CUU CG-3′
RNA^{TPT3_RT}	5′-GGG UUA ACU UUA AGA AGG AGA AAA ACA UGU ACA AAA AGU ACA AGA r**TPT3**GG ACU ACA AGG ACG ACG ACG ACA AGU AAG CUU CG-3′

6 References

[1] Crick, F., Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. https://doi.org/10.1038/227561a0.

[2] Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA Dodecamer: Conformation and Dynamics. *Proc. Natl. Acad. Sci.* **1981**, *78* (4), 2179–2183. https://doi.org/10.1073/pnas.78.4.2179.

[3] Sekine, S. I.; Nureki, O.; Shimada, A.; Vassylyev, D. G.; Yokoyama, S., Structural Basis for Anticodon Recognition by Discriminating Glutamyl-TRNA Synthetase. *Nat. Struct. Biol.* **2001**, *8* (3), 203–206. https://doi.org/10.1038/84927.

[4] Niederhut, M. S., Three-Dimensional Structures of the Three Human Class I Alcohol Dehydrogenases. *Protein Sci.* **2001**, *10* (4), 697–706. https://doi.org/10.1110/ps.45001.

[5] Lingner, J., Reverse Transcriptase Motifs in the Catalytic Subunit of Telomerase. *Science* **1997**, *276* (5312), 561–567. https://doi.org/10.1126/science.276.5312.561.

[6] Finnegan, D. J., Eukaryotic Transposable Elements and Genome Evolution. *Trends Genet.* **1989**, *5* (4), 103–107. https://doi.org/10.1016/0168-9525(89)90039-5.

[7] Teng, S.-C.; Kim, B.; Gabriel, A., Retrotransposon Reverse-Transcriptase-Mediated Repair of Chromosomal Breaks. *Nature* **1996**, *383* (6601), 641–644. https://doi.org/10.1038/383641a0.

[8] Inouye, S.; Inouye, M., Structure, Function, and Evolution of Bacterial Reverse Transcriptase. *Virus Genes* **1995**, *11* (2–3), 81–94. https://doi.org/10.1007/BF01728650.

[9] Inouye, S.; Hsu, M.-Y.; Eagle, S.; Inouye, M., Reverse Transcriptase Associated with the Biosynthesis of the Branched RNA-Linked MsDNA in Myxococcus Xanthus. *Cell* **1989**, *56* (4), 709–717. https://doi.org/10.1016/0092-8674(89)90593-X.

[10] Ma, W., What Does “the RNA World” Mean to “the Origin of Life”? *Life* **2017**, *7* (4), 49. https://doi.org/10.3390/life7040049.

[11] Das, S. R.; Fong, R.; Piccirilli, J., A. Nucleotide Analogues to Investigate RNA Structure and Function. *Curr. Opin. Chem. Biol.* **2005**, *9*, 585–593. https://doi.org/10.1016/j.cbpa.2005.10.009.

[12] Hannon, G. J.; Rivas, F. V.; Murchison, E. P.; Steitz, J. A., The Expanding Universe of Noncoding RNAs. *Cold Spring Harb. Symp. Quant. Biol.* **2006**, *71*, 551–564. https://doi.org/10.1101/sqb.2006.71.064.

[13] Kubota, M.; Tran, C.; Spitale, R. C., Progress and Challenges for Chemical Probing of RNA Structure inside Living Cells. *Nat. Chem. Biol.* **2015**, *11* (12), 933–941. https://doi.org/10.1038/nchembio.1958.

[14] Vázquez-Salazar, A.; Lazcano, A., Early Life: Embracing the RNA World. *Curr. Biol.* **2018**, *28* (5), R220–R222. https://doi.org/10.1016/j.cub.2018.01.055.

[15] Larsson, A.; Reichard, P., Enzymatic Synthesis of Deoxyribonucleotides X. Reduction of Purine Ribonucleotides; Allosteric Behavior and Substrate Specificity of the Enzyme System from Escherichia Coli B. *J. Biol. Chem.* **1966**, *241*, 2540–2549.

[16] Bernhardt, H. S., The RNA World Hypothesis: The Worst Theory of the Early Evolution of Life (except for All the Others)A. *Biol. Direct* **2012**, *7* (1), 23. https://doi.org/10.1186/1745-6150-7-23.

[17] Higgs, P. G.; Lehman, N., The RNA World: Molecular Cooperation at the Origins of Life. *Nat. Rev. Genet.* **2015**, *16* (1), 7–17. https://doi.org/10.1038/nrg3841.

[18] Orgel, L. E., Some Consequences of the RNA World Hypothesis. *Orig. Life Evol. Biosph.* **2003**, *33* (2), 211–218. https://doi.org/10.1023/A:1024616317965.

[19] Joyce, G. F., The Antiquity of RNA-Based Evolution. *Nature* **2002**, *418*, 214–221. https://doi.org/10.1038/418214a.

[20] Groen, J.; Deamer, D. W.; Kros, A.; Ehrenfreund, P., Polycyclic Aromatic Hydrocarbons as Plausible Prebiotic Membrane Components. *Orig. Life Evol. Biosph.* **2012**, *42* (4), 295–306. https://doi.org/10.1007/s11084-012-9292-3.

[21] Walter, G., The RNA World. *Nature* **1986**, *319*, 86.

[22] Dworkin, J. P.; Lazcano, A.; Miller, S. L., The Roads to and from the RNA World. *J. Theor. Biol.* **2003**, *222*, 127–134. https://doi.org/10.1016/S0022-5193(03)00020-1.
[23] Nelson, K. E.; Levy, M.; Miller, S. L., Peptide Nucleic Acids Rather than RNA May Have Been the First Genetic Molecule. *Proc. Natl. Acad. Sci.* **2000**, *97* (8), 3868–3871. https://doi.org/10.1073/pnas.97.8.3868.
[24] Zhang, L.; Peritz, A.; Meggers, E. A., Simple Glycol Nucleic Acid. *J. Am. Chem. Soc.* **2005**, *127* (12), 4174–4175. https://doi.org/10.1021/ja042564z.
[25] Eschenmoser, A., The TNA-Family of Nucleic Acid Systems: Properties and Prospects. *Orig. Life Evol. Biosph.* **2004**, *34* (3), 277–306. https://doi.org/10.1023/B:ORIG.0000016450.59665.f4.
[26] Lazcano, A.; Bada, J. L., The 1953 Stanley L. Miller Experiment: Fifty Years of Prebiotic Organic Chemistry. *Orig. Life Evol. Biosph.* **2003**, *33* (3), 235–242. https://doi.org/10.1023/A:1024807125069.
[27] Ferus, M.; Pietrucci, F.; Saitta, A. M.; Knížek, A.; Kubelík, P.; Ivanek, O.; Shestivska, V.; Civiš, S., Formation of Nucleobases in a Miller–Urey Reducing Atmosphere. *Proc. Natl. Acad. Sci.* **2017**, *114* (17), 4306–4311. https://doi.org/10.1073/pnas.1700010114.
[28] Bada, J. L., Prebiotic Soup - Revisiting the Miller Experiment. *Science* **2003**, *300* (5620), 745–746. https://doi.org/10.1126/science.1085145.
[29] Benner, S. A.; Kim, H.-J.; Carrigan, M. A., Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA. *Acc. Chem. Res.* **2012**, *45* (12), 2025–2034. https://doi.org/10.1021/ar200332w.
[30] Pressman, A.; Blanco, C.; Chen, I. A. ,The RNA World as a Model System to Study the Origin of Life. *Curr. Biol.* **2015**, *25* (19), R953–R963. https://doi.org/10.1016/j.cub.2015.06.016.
[31] Powner, M. W.; Gerland, B.; Sutherland, J. D., Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions. *Nature* **2009**, *459* (7244), 239–242. https://doi.org/10.1038/nature08013.
[32] Pearce, B. K. D.; Pudritz, R. E.; Semenov, D. A.; Henning, T. K., Origin of the RNA World: The Fate of Nucleobases in Warm Little Ponds. *Proc. Natl. Acad. Sci.* **2017**, *114* (43), 11327–11332. https://doi.org/10.1073/pnas.1710339114.
[33] Cech, T. R., A Model for the RNA-Catalyzed Replication of RNA. *Proc. Natl. Acad. Sci.* **1986**, *83* (12), 4360–4363. https://doi.org/10.1073/pnas.83.12.4360.
[34] Robertson, M. P.; Joyce, G. F., The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (5), a003608–a003608. https://doi.org/10.1101/cshperspect.a003608.
[35] Cheng, L. K. L.; Unrau, P. J., Closing the Circle: Replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (10), a002204–a002204. https://doi.org/10.1101/cshperspect.a002204.
[36] Robertson, M. P.; Scott, W. G., The Structural Basis of Ribozyme-Catalyzed RNA Assembly. *Science (80-.).* **2007**, *315* (5818), 1549–1553. https://doi.org/10.1126/science.1136231.
[37] Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E., The Case for an Ancestral Genetic System Involving Simple Analogues of the Nucleotides. *Proc. Natl. Acad. Sci.* **1987**, *84* (13), 4398–4402. https://doi.org/10.1073/pnas.84.13.4398.
[38] Li, L.; Francklyn, C.; Carter, C. W., Aminoacylating Urzymes Challenge the RNA World Hypothesis. *J. Biol. Chem.* **2013**, *288* (37), 26856–26863. https://doi.org/10.1074/jbc.M113.496125.
[39] Thomas, B.; Akoulitchev, A. V., Mass Spectrometry of RNA. *Trends Biochem. Sci.* **2006**, *31* (3), 173–181. https://doi.org/10.1016/j.tibs.2006.01.004.
[40] Mattick, J. S.; Makunin, I. V., Non-Coding RNA. *Hum. Mol. Genet.* **2006**, *15* (suppl_1), R17–R29. https://doi.org/10.1093/hmg/ddl046.
[41] Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M. F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B. W.; Cassady, J. P.; et al., Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals. *Nature* **2009**, *458* (7235), 223–227. https://doi.org/10.1038/nature07672.

[42] Khachane, A. N.; Harrison, P. M., Mining Mammalian Transcript Data for Functional Long Non-Coding RNAs. *PLoS One* **2010**, *5* (4), e10316. https://doi.org/10.1371/journal.pone.0010316.
[43] Eddy, S. R., Non–Coding RNA Genes and the Modern RNA World. *Nat. Rev. Genet.* **2001**, *2*, 919–929. https://doi.org/10.1038/35103511.
[44] Geisler, S.; Coller, J., RNA in Unexpected Places: Long Non-Coding RNA Functions in Diverse Cellular Contexts. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (11), 699–712. https://doi.org/10.1038/nrm3679.
[45] Tang, G., SiRNA and MiRNA: An Insight into RISCs. *Trends Biochem. Sci.* **2005**, *30* (2), 106–114. https://doi.org/10.1016/j.tibs.2004.12.007.
[46] Carthew, R. W.; Sontheimer, E. J., Origins and Mechanisms of MiRNAs and SiRNAs. *Cell* **2009**, *136*, 642–655. https://doi.org/10.1016/j.cell.2009.01.035.
[47] Bartel, D. P., MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **2009**, *136*, 215–233. https://doi.org/10.1016/j.cell.2009.01.002.
[48] Hartmann, R. K.; Bindereif, A.; Schön, A.; Westhof, E., *Handbook of RNA Biochemistry*, 2nd ed.; Hartmann, R. K., Bindereif, A., Schön, A., Westhof, E., Eds.; Wiley-VCH: Weinheim, **2014**.
[49] Srivastava, S. K.; Arora, S.; Singh, S.; Bhardwaj, A.; Averett, C.; Singh, A. P., MicroRNAs in Pancreatic Malignancy: Progress and Promises. *Cancer Lett.* **2014**, *347*, 167–174. https://doi.org/10.1016/j.canlet.2014.02.015.
[50] Tanner, N. K., Ribozymes: The Characteristics and Properties of Catalytic RNAs. *FEMS Microbiol. Rev.* **1999**, *23*, 257–275.
[51] Hartig, J. S., *Ribozymes*; 2012. https://doi.org/10.1007/978-1-62703-239-1_1.
[52] Doudna, J. A.; Lorsch, J. R., Ribozyme Catalysis: Not Different, Just Worse. *Nat. Struct. Mol. Biol.* **2005**, *12* (5), 395–402. https://doi.org/10.1038/nsmb932.
[53] Tang, J.; Breaker, R. R., Structural Diversity of Self-Cleaving Ribozymes. *PNAS* **2000**, *97* (11), 5784–5789. https://doi.org/10.1073/pnas.97.11.5784.
[54] Ferré-D'Amaré, A. R.; Scott, W. G., Small Self-Cleaving Ribozymes. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, 1–10. https://doi.org/10.1101/cshperspect.a003574.
[55] Webb, C.-H. T.; Lupták, A., HDV-like Self-Cleaving Ribozymes. *RNA Biol.* **2011**, *8* (5), 719–727. https://doi.org/10.4161/rna.8.5.16226.
[56] Gibb, E. A.; Brown, C. J.; Lam, W. L., The Functional Role of Long Non-Coding RNA in Human Carcinomas. *Mol. Cancer* **2011**, *10* (1), 38. https://doi.org/10.1186/1476-4598-10-38.
[57] Ponjavic, J.; Oliver, P. L.; Lunter, G.; Ponting, C. P., Genomic and Transcriptional Co-Localization of Protein-Coding and Long Non-Coding RNA Pairs in the Developing Brain. *PLoS Genet.* **2009**, *5* (8), e1000617. https://doi.org/10.1371/journal.pgen.1000617.
[58] Ransohoff, J. D.; Wei, Y.; Khavari, P. A., The Functions and Unique Features of Long Intergenic Non-Coding RNA. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (3), 143–157. https://doi.org/10.1038/nrm.2017.104.
[59] Sun, J.; Chen, X.; Wang, Z.; Guo, M.; Shi, H.; Wang, X.; Cheng, L.; Zhou, M. A., Potential Prognostic Long Non-Coding RNA Signature to Predict Metastasis-Free Survival of Breast Cancer Patients. *Sci. Rep.* **2015**, *5*. https://doi.org/10.1038/srep16553.
[60] Anastasiadou, E.; Jacob, L. S.; Slack, F. J., Non-Coding RNA Networks in Cancer. *Nat. Rev. Cancer* **2018**, *18*, 5–18. https://doi.org/10.1038/nrc.2017.99.
[61] Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R., Control of Gene Expression by a Natural Metabolite-Responsive Ribozyme. *Nature* **2004**, *428*, 281–286. https://doi.org/10.1038/nature02362.
[62] McCown, P. J.; Roth, A.; Breaker, R. R., An Expanded Collection and Refined Consensus Model of GlmS Ribozymes. *RNA* **2011**, *17*, 728–736. https://doi.org/10.1261/rna.2590811.

[63] Schüller, A.; Matzner, D.; Lünse, C. E.; Wittmann, V.; Schumacher, C.; Unsleber, S.; Brötz-Oesterhelt, H.; Mayer, C.; Bierbaum, G.; Mayer, G., Activation of the GlmS Ribozyme Confers Bacterial Growth Inhibition. *ChemBioChem* **2017**, *18* (5), 435–440. https://doi.org/10.1002/cbic.201600491.

[64] Serganov, A.; Nudler, E. A., Decade of Riboswitches. *Cell* **2013**, *152*, 17–24. https://doi.org/10.1016/j.cell.2012.12.024.

[65] McCarthy, T. J.; Plog, M. A.; Floy, S. A.; Jansen, J. A.; Soukup, J. K.; Soukup, G. A., Ligand Requirements for GlmS Ribozyme Self-Cleavage. *Chem. Biol.* **2005**, *12*, 1221–1226. https://doi.org/10.1016/j.chembiol.2005.09.006.

[66] Soukup, G. A., Core Requirements for GlmS Ribozyme Self-Cleavage Reveal a Putative Pseudoknot Structure. *Nucleic Acids Res.* **2006**, *34* (3), 968–975. https://doi.org/10.1093/nar/gkj497.

[67] Wilkinson, S. R.; Been, M. D., A Pseudoknot in the 3' Non-Core Region of the GlmS Ribozyme Enhances Self-Cleavage Activity. *RNA* **2005**, *11*, 1788–1794. https://doi.org/10.1261/rna.2203605.

[68] Hampel, K. J.; Tinsley, M. M., Evidence for Preorganization of the GlmS Ribozyme Ligand Binding Pocket †. *Biochemistry* **2006**, *45* (25), 7861–7871. https://doi.org/10.1021/bi060337z.

[69] Zhang, S.; Ganguly, A.; Goyal, P.; Bingaman, J. L.; Bevilacqua, P. C.; Hammes-Schiffer, S., Role of the Active Site Guanine in the GlmS Ribozyme Self-Cleavage Mechanism: Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *J. Am. Chem. Soc.* **2015**, *137* (2), 784–798. https://doi.org/10.1021/ja510387y.

[70] Klein, D. J.; Ferré-D'Amaré, A. R., Structural Basis of GlmS Ribozyme Activation by Glucosamine-6-Phosphate. *Science* **2006**, *313*, 1752–1756. https://doi.org/10.1126/science.1129666.

[71] Klein, D. J.; Wilkinson, S. R.; Been, M. D.; Ferré-D'Amaré, A. R., Requirement of Helix P2.2 and Nucleotide G1 for Positioning the Cleavage Site and Cofactor of the GlmS Ribozyme. *J. Mol. Biol.* **2007**, *373*, 178–189. https://doi.org/c.

[72] Abera, G.; Bernahu, G.; Tekewe, A., Ribozymes: Nucleic Acid Enzymes with Potential Pharmaceutical Applications - A Review. *Pharmacophore* **2012**, *3* (3), 164–178.

[73] Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A., Structural Investigation of the GlmS Ribozyme Bound to Its Catalytic Cofactor. *Chem. Biol.* **2007**, *14*, 97–105. https://doi.org/10.1016/j.chembiol.2006.12.005

[74] Cochrane, J. C.; Lipchock, S. V; Smith, K. D.; Strobel, S. A., Structural and Chemical Basis for Glucosamine 6-Phosphate Binding and Activation of the GlmS Ribozyme. *Biochemistry* **2009**, *48* (15), 3239–3246. https://doi.org/10.1021/bi802069p.

[75] Brockdorff, N.; Ashworth, A.; Kay, G. F.; Cooper, P.; Smith, S.; McCabe, V. M.; Norris, D. P.; Penny, G. D.; Patel, D.; Rastan, S., Conservation of Position and Exclusive Expression of Mouse Xist from the Inactive X Chromosome. *Nature* **1991**, *351* (6324), 329–331. https://doi.org/10.1038/351329a0.

[76] Brockdorff, N.; Ashworth, A.; Kay, G. F.; McCabe, V. M.; Norris, D. P.; Cooper, P. J.; Swift, S.; Rastan, S., The Product of the Mouse Xist Gene Is a 15 Kb Inactive X-Specific Transcript Containing No Conserved ORF and Located in the Nucleus. *Cell* **1992**, *71* (3), 515–526. https://doi.org/10.1016/0092-8674(92)90519-I.

[77] Brown, C. J.; Hendrich, B. D.; Rupert, J. L.; Lafrenière, R. G.; Xing, Y.; Lawrence, J.; Willard, H. F., The Human XIST Gene: Analysis of a 17 Kb Inactive X-Specific RNA That Contains Conserved Repeats and Is Highly Localized within the Nucleus. *Cell* **1992**, *71* (3), 527–542. https://doi.org/10.1016/0092-8674(92)90520-M.

[78] Panning, B.; Dausman, J.; Jaenisch, R. X., Chromosome Inactivation Is Mediated by Xist RNA Stabilization. *Cell* **1997**, *90* (5), 907–916. https://doi.org/10.1016/S0092-8674(00)80355-4.

[79] Smola, M. J.; Christy, T. W.; Inoue, K.; Nicholson, C. O.; Friedersdorf, M.; Keene, J. D.; Lee, D. M.; Calabrese, J. M.; Weeks, K. M., SHAPE Reveals Transcript-Wide Interactions, Complex Structural Domains, and Protein Interactions across the Xist LncRNA in Living Cells. *Proc. Natl. Acad. Sci.* **2016**, *113* (37), 10322–10327. https://doi.org/10.1073/pnas.1600008113.

[80] Maenner, S.; Blaud, M.; Fouillen, L.; Savoye, A.; Marchand, V.; Dubois, A.; Sanglier-Cianférani, S.; Van Dorsselaer, A.; Clerc, P.; Avner, P.; et al., 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. *PLoS Biol.* **2010**, *8* (1), e1000276. https://doi.org/10.1371/journal.pbio.1000276.

[81] Fang, R.; Moss, W. N.; Rutenberg-Schoenberg, M.; Simon, M. D., Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLOS Genet.* **2015**, *11* (12), e1005668. https://doi.org/10.1371/journal.pgen.1005668.

[82] Gilbert, S. L.; Pehrson, J. R.; Sharp, P. A., XIST RNA Associates with Specific Regions of the Inactive X Chromatin. *J. Biol. Chem.* **2000**, *275* (47), 36491–36494. https://doi.org/10.1074/jbc.C000409200.

[83] Plath, K.; Mlynarczyk-Evans, S.; Nusinow, D. A.; Panning, B., Xist RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* **2002**, *36* (1), 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433.

[84] Pontier, D. B.; Gribnau, J., Xist Regulation and Function EXplored. *Hum. Genet.* **2011**, *130* (2), 223–236. https://doi.org/10.1007/s00439-011-1008-7.

[85] Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A., Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **1998**, *395* (6702), 567–574. https://doi.org/10.1038/26912.

[86] Fedor, M. J.; Williamson, J. R., The Catalytic Diversity of RNAs. *Nat. Rev. - Mol. cell Biol.* **2005**, *6*, 399–412. https://doi.org/10.1038/nrm1647.

[87] Ke, A.; Doudna, J. A., Crystallization of RNA and RNA-Protein Complexes. *Methods* **2004**, *34* (3), 408–414. https://doi.org/10.1016/j.ymeth.2004.03.027.

[88] Wang, J.; Smerdon, S. J.; Jager, J.; Kohlstaedt, L. A.; Rice, P. A.; Friedman, J. M.; Steitz, T. A., Structural Basis of Asymmetry in the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Heterodimer. *Proc. Natl. Acad. Sci.* **1994**, *91* (15), 7242–7246. https://doi.org/10.1073/pnas.91.15.7242.

[89] Grosshans, C. A.; Cech, T. R. A., Hammerhead Ribozyme Allows Synthesis of a New Form of the Tetrahymena Ribozyme Homogeneous in Length with a 3' End Blocked for Transesterification. *Nucleic Acids Res.* **1991**, *19* (14), 3875–3880. https://doi.org/10.1093/nar/19.14.3875.

[90] Wahl, M. C.; Ramakrishnan, B.; Ban, C.; Chen, X.; Sundaralingam, M., RNA – Synthesis, Purification and Crystallization. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1996**, *52* (4), 668–675. https://doi.org/10.1107/S0907444996002788.

[91] Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A., A General Module for RNA Crystallization. *J. Mol. Biol.* **1998**, *279* (3), 621–631. https://doi.org/10.1006/jmbi.1998.1789.

[92] Krissinel, E.; Henrick, K., Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **2007**, *372* (3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

[93] Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; et al., The Protein Data Bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2002**, *58* (6), 899–907. https://doi.org/10.1107/S0907444902003451.

[94] Varani, G.; Aboul-ela, F.; Allain, F. H.-T., NMR Investigation of RNA Structure. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29* (1–2), 51–127. https://doi.org/10.1016/0079-6565(96)01028-X.

[95] Kulikov, K.; Nozinovic, S.; Kath-Schorr, S., The 5′-AG5CC-3′ Fragment from the Human CPEB3 Ribozyme Forms an Ultrastable Parallel RNA G-Quadruplex. *ChemBioChem* **2017**, *18* (11), 969–973. https://doi.org/10.1002/cbic.201700028.

[96] Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H., NMR Spectroscopy of RNA. *ChemBioChem* **2003**, *4* (10), 936–962. https://doi.org/10.1002/cbic.200300700.

[97] Karplus, M.; McCammon, J. A., Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652. https://doi.org/10.1038/nsb0902-646.

[98] Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; et al., RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chem. Rev.* **2018**, *118* (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.

[99] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.
[100] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.
[101] Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32* (9), 1929–1943. https://doi.org/10.1002/jcc.21777.
[102] Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E., RNA Force Field with Accuracy Comparable to State-of-the-Art Protein Force Fields. *Proc. Natl. Acad. Sci.* **2018**, *115* (7), E1346–E1355. https://doi.org/10.1073/pnas.1713027115.
[103] Leamy, K. A.; Assmann, S. M.; Mathews, D. H.; Bevilacqua, P. C., Bridging the Gap between in Vitro and in Vivo RNA Folding. *Q. Rev. Biophys.* **2016**, *49*, e10. https://doi.org/10.1017/S003358351600007X.
[104] Porecha, R.; Herschlag, D., RNA Radiolabeling. *Methods Enzymol.* **2013**, *530*, 255–279. https://doi.org/10.1016/B978-0-12-420037-1.00014-2.
[105] Höltke, H.-J.; Kessler, C., Non-Radioactive Labeling of RNA Transcripts in Vitro with the Hapten Digoxigenin (DIG); Hybridization and ELISA-Based Detection. *Nucleic Acids Res.* **1990**, *18* (19), 5843–5851. https://doi.org/10.1093/nar/18.19.5843.
[106] Beckert, B.; Benoît, M. *RNA*; Nielsen, H., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2011; Vol. 703. https://doi.org/10.1007/978-1-59745-248-9.
[107] Levsky, J. M., Fluorescence in Situ Hybridization: Past, Present and Future. *J. Cell Sci.* **2003**, *116* (14), 2833–2838. https://doi.org/10.1242/jcs.00633.
[108) Tan, W.; Wang, K.; Drake, T. J., Molecular Beacons. *Curr. Opin. Chem. Biol.* **2004**, *8*, 547–553. https://doi.org/10.1016/j.cbpa.2004.08.010.
[109] Jockusch, S.; Mart?, A. A.; Turro, N. J.; Li, Z.; Li, X.; Ju, J.; Stevens, N.; Akins, D. L., Spectroscopic Investigation of a FRET Molecular Beacon Containing Two Fluorophores for Probing DNA/RNA Sequences. *Photochem. Photobiol. Sci.* **2006**, *5* (5), 493. https://doi.org/10.1039/b600213g.
[110] Santangelo, P. J.; Nix, B.; Tsourkas, A.; Bao, G., Dual FRET Molecular Beacons for MRNA Detection in Living Cells. *Nucleic Acids Res.* **2004**, *32* (6), e57. https://doi.org/10.1093/nar/gnh062.
[111] Jares-Erijman, E. A.; Jovin, T. M., FRET Imaging. *Nat. Biotechnol.* **2003**, *21* (11), 1387–1395. https://doi.org/10.1038/nbt896.
[112] Roy, R.; Hohng, S.; Ha, T. A., Practical Guide to Single-Molecule FRET. *Nat. Methods* **2008**, *5* (6), 507–516. https://doi.org/10.1038/nmeth.1208.
[113] Qin, P. Z.; Pyle, A. M., Site-Specific Labeling of RNA with Fluorophores and Other Structural Probes. *Methods* **1999**, *18*, 60–70. https://doi.org/10.1006/meth.1999.0757.
[114] Solomatin, S. V.; Greenfeld, M.; Chu, S.; Herschlag, D., Multiple Native States Reveal Persistent Ruggedness of an RNA Folding Landscape. *Nature* **2010**, *463* (7281), 681–684. https://doi.org/10.1038/nature08717.
[115] Bood, M.; Sarangamath, S.; Wranne, M. S.; Grøtli, M.; Wilhelmsson, L. M., Fluorescent Nucleobase Analogues for Base–Base FRET in Nucleic Acids: Synthesis, Photophysics and Applications. *Beilstein J. Org. Chem.* **2018**, *14*, 114–129. https://doi.org/10.3762/bjoc.14.7.
[116] Sexton, A. N.; Wang, P. Y.; Rutenberg-Schoenberg, M.; Simon, M. D., Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry* **2017**, *56* (35), 4713–4721. https://doi.org/10.1021/acs.biochem.7b00323.
[117] Hajdin, C. E.; Bellaousov, S.; Huggins, W.; Leonard, C. W.; Mathews, D. H.; Weeks, K. M., Accurate SHAPE-Directed RNA Secondary Structure Modeling, Including Pseudoknots. *Proc. Natl. Acad. Sci.* **2013**, *110* (14), 5498–5503. https://doi.org/10.1073/pnas.1219988110.

[118] Brunel, C.; Romby, P., Probing RNA Structure and RNA-Ligand Complexes with Chemical Probes. *Methods Enzymol.* **2000**, *318*.
[119] Alm, E. W.; Oerther, D. B.; Larsen, N.; Stahl, D. A.; Raskin, L., The Oligonucleotide Probe Database. *Appl. Environ. Microbiol.* **1996**, *62* (10), 3557–3559.
[120] Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J.-P.; Ehresmann, B., Probing the Structure of RNAs in Solution. *Nucleic Acids Res.* **1987**, *15* (22), 9109–9128. https://doi.org/10.1093/nar/15.22.9109.
[121] Moazed, D.; Stern, S.; Noller, H. F., Rapid Chemical Probing of Conformation in 16 S Ribosomal RNA and 30 S Ribosomal Subunits Using Primer Extension. *J. Mol. Biol.* **1986**, *187* (3), 399–416. https://doi.org/10.1016/0022-2836(86)90441-9.
[122] Paredes, E.; Evans, M.; Das, S. R., RNA Labeling, Conjugation and Ligation. *Methods* **2011**, *54*, 251–259. https://doi.org/10.1016/j.ymeth.2011.02.008.
[123] Piton, N.; Mu, Y.; Stock, G.; Prisner, T. F.; Schiemann, O.; Engels, J. W., Base-Specific Spin-Labeling of RNA for Structure Determination. *Nucleic Acids Res.* **2007**, *35* (9), 3128–3143. https://doi.org/10.1093/nar/gkm169.
[124] Defrancq, E.; Messaoudi, S., Palladium-Mediated Labeling of Nucleic Acids. *ChemBioChem* **2017**, *18* (5), 426–431. https://doi.org/10.1002/cbic.201600599.
[125] Best, M. D., Click Chemistry and Bioorthogonal Reactions: Unprecedented Selectivity in the Labeling of Biological Molecules. *Biochemistry* **2009**, *48* (28), 6571–6584. https://doi.org/10.1021/bi9007726.
[126] Merkel, M.; Peewasan, K.; Arndt, S.; Ploschik, D.; Wagenknecht, H.-A., Copper-Free Postsynthetic Labeling of Nucleic Acids by Means of Bioorthogonal Reactions. *ChemBioChem* **2015**, *16*, 1541–1553. https://doi.org/10.1002/cbic.201500199.
[127] Kerzhner, M.; Abdullin, D.; Więcek, J.; Matsuoka, H.; Hagelueken, G.; Schiemann, O.; Famulok, M., Post-Synthetic Spin-Labeling of RNA through Click Chemistry for PELDOR Measurements. *Chem. - A Eur. J.* **2016**, *22*, 12113–12121. https://doi.org/10.1002/chem.201601897.
[128] Samanta, B.; Seikowski, J.; Höbartner, C., Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA. *Angew. Chemie Int. Ed.* **2016**, *55* (5), 1912–1916. https://doi.org/10.1002/anie.201508893.
[129] Walunj, M. B.; Tanpure, A. A.; Srivatsan, S. G., Post-Transcriptional Labeling by Using Suzuki–Miyaura Cross-Coupling Generates Functional RNA Probes. *Nucleic Acids Res.* **2018**, *46* (11), e65–e65. https://doi.org/10.1093/nar/gky185.
[130] Srivatsan, S. G.; Tor, Y., Fluorescent Pyrimidine Ribonucleotide: Synthesis, Enzymatic Incorporation, and Utilization. *J. Am. Chem. Soc.* **2007**, *129* (7), 2044–2053. https://doi.org/10.1021/ja066455r.
[131] Kore, A. R.; Senthilvelan, A.; Shanmugasundaram, M., Highly Chemoselective Palladium-Catalyzed Sonogashira Coupling of 5-Iodouridine-5′-Triphosphates with Propargylamine: A New Efficient Method for the Synthesis of 5-Aminopropargyl-Uridine-5′-Triphosphates. *Tetrahedron Lett.* **2012**, *53*, 3070–3072. https://doi.org/10.1016/j.tetlet.2012.04.023.
[132] Motorin, Y.; Burhenne, J.; Teimer, R.; Koynov, K.; Willnow, S.; Weinhold, E.; Helm, M., Expanding the Chemical Scope of RNA: Methyltransferases to Site-Specific Alkynylation of RNA for Click Labeling. *Nucleic Acids Res.* **2011**, *39* (5), 1943–1952. https://doi.org/10.1093/nar/gkq825.
[133] Anhäuser, L.; Rentmeister, A., Enzyme-Mediated Tagging of RNA. *Curr. Opin. Biotechnol.* **2017**, *48*, 69–76. https://doi.org/10.1016/j.copbio.2017.03.013.
[134] Switzer, C.; Moroney, S. E.; Benner, S. A., Enzymatic Incorporation of a New Base Pair into DNA and RNA. *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323. https://doi.org/10.1038/343033a0.
[135] Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A., Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* **1990**, *343*, 33–37. https://doi.org/10.1038/343033a0.
[136] Switzer, C. Y.; Moroney, S. E.; Benner, S. A., Enzymic Recognition of the Base Pair Between Isocytidine and Isoguanosine. *Biochemistry* **1993**, *32*, 10489–10496. https://doi.org/10.1021/bi00090a027.

[137] Moran, S.; Ren, R. X. F.; Rumney; Kool, E. T., Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication. *J. Am. Chem. Soc.* **1997**, *119* (8), 2056–2057. https://doi.org/10.1021/ja963718g.
[138] Morales, J. C.; Kool, E. T., Efficient Replication between Non-Hydrogen-Bonded Nucleoside Shape Analogs. *Nat. Struct. Biol.* **1998**, *5* (11), 950–954. https://doi.org/10.1038/2925.
[139] McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E., Efforts toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base. *J. Am. Chem. Soc.* **1999**, *121* (49), 11585–11586. https://doi.org/10.1021/ja9925150.
[140] Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E., Efforts toward the Expansion of the Genetic Alphabet: Information Storage and Replication with Unnatural Hydrophobic Base Pairs. *J. Am. Chem. Soc.* **2000**, *122*, 3274–3287. https://doi.org/10.1021/ja9940064.
[141] Martinot, T. A.; Benner, S. A., Artificial Genetic Systems: Exploiting the “Aromaticity” Formalism To Improve the Tautomeric Ratio for Isoguanosine Derivatives. *J. Org. Chem.* **2004**, *69* (11), 3972–3975. https://doi.org/10.1021/jo0497959.
[142] Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Fajardo, D.; Moussatche, P., Alternative Watson-Crick Synthetic Genetic Systems. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (11). https://doi.org/10.1101/cshperspect.a023770.
[143] Huttert, D.; Benner, S. A., Expanding the Genetic Alphabet: Non-Epimerizing Nucleoside with the PyDDA Hydrogen-Bonding Pattern. *J. Org. Chem.* **2003**, *68* (25), 9839–9842. https://doi.org/10.1021/jo034900k.
[144] Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A., Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* **2011**, *133*, 15105–15112. https://doi.org/10.1021/ja204910n.
[145] Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A., Artificially Expanded Genetic Information System: A New Base Pair with an Alternative Hydrogen Bonding Pattern. *Nucleic Acids Res.* **2006**, *34* (21), 6095–6101. https://doi.org/10.1093/nar/gkl633.
[146] Kim, H.-J.; Leal, N. A.; Benner, S. A., 2′-Deoxy-1-Methylpseudocytidine, a Stable Analog of 2′-Deoxy-5-Methylisocytidine. *Bioorg. Med. Chem.* **2009**, *17* (10), 3728–3732. https://doi.org/10.1016/j.bmc.2009.03.047.
[147] Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; et al., Hachimoji DNA and RNA: A Genetic System with Eight Building Blocks. *Science* **2019**, *363* (6429), 884–887. https://doi.org/10.1126/science.aat0971.
[148] Kim, H.-J.; Leal, N. A.; Hoshika, S.; Benner, S. A., Ribonucleosides for an Artificially Expanded Genetic Information System. *J. Org. Chem.* **2014**, *79* (7), 3194–3199. https://doi.org/10.1021/jo402665d.
[149] Biondi, E.; Lane, J. D.; Das, D.; Dasgupta, S.; Piccirilli, J. A.; Hoshika, S.; Bradley, K. M.; Krantz, B. A.; Benner, S. A., Laboratory Evolution of Artificially Expanded DNA Gives Redesignable Aptamers That Target the Toxic Form of Anthrax Protective Antigen. *Nucleic Acids Res.* **2016**, *44* (20), gkw890. https://doi.org/10.1093/nar/gkw890.
[150] Wu, Y. X.; Kwon, Y. J., Aptamers: The “Evolution” of SELEX. *Methods* **2016**, *106*, 21–28. https://doi.org/10.1016/j.ymeth.2016.04.020.
[151] Leal, N. A.; Kim, H.-J.; Hoshika, S.; Kim, M.-J.; Carrigan, M. A.; Benner, S. A., Transcription, Reverse Transcription, and Analysis of RNA Containing Artificial Genetic Components. *ACS Synth. Biol.* **2015**, *4* (4), 407–413. https://doi.org/10.1021/sb500268n.
[152] Benner, S. A.; Sismour, A. M., Synthetic Biology. *Nat. Rev. Genet.* **2005**, *6* (7), 533–543. https://doi.org/10.1038/nrg1637.

[153] Georgiadis, M. M.; Singh, I.; Kellett, W. F.; Hoshika, S.; Benner, S. A.; Richards, N. G. J., Structural Basis for a Six Nucleotide Genetic Alphabet. *J. Am. Chem. Soc.* **2015**, *137*, 6947–6955. https://doi.org/10.1021/jacs.5b03482.

[154] Singh, I.; Laos, R.; Hoshika, S.; Benner, S. A.; Georgiadis, M. M., Snapshots of an Evolved DNA Polymerase Pre- and Post-Incorporation of an Unnatural Nucleotide. *Nucleic Acids Res.* **2018**, *46* (15), 7977–7988. https://doi.org/10.1093/nar/gky552.

[155] Matsuura, M. F.; Shaw, R. W.; Moses, J. D.; Kim, H.-J.; Kim, M.-J.; Kim, M.-S.; Hoshika, S.; Karalkar, N.; Benner, S. A., Assays To Detect the Formation of Triphosphates of Unnatural Nucleotides: Application to Escherichia Coli Nucleoside Diphosphate Kinase. *ACS Synth. Biol.* **2016**, *5* (3), 234–240. https://doi.org/10.1021/acssynbio.5b00172.

[156] Chen, F.; Zhang, Y.; Daugherty, A. B.; Yang, Z.; Shaw, R.; Dong, M.; Lutz, S.; Benner, S. A., Biological Phosphorylation of an Unnatural Base Pair (UBP) Using a Drosophila Melanogaster Deoxynucleoside Kinase (DmdNK) Mutant. *PLoS One* **2017**, *12* (3), 1–10. https://doi.org/10.1371/journal.pone.0174163.

[157] Ishikawa, M.; Hirao, I.; Yokoyama, S., Synthesis of 3-(2-Deoxy-β-D-Ribofuranosyl)Pyridin-2-One and 2-Amino-6- (N,N-Dimethylamino)-9-(2-Deoxy-β-D-Ribofuranosyl)Purine Derivatives for an Unnatural Base Pair. *Tetrahedron Lett.* **2000**, *41*, 3931–3934. https://doi.org/10.1016/S0040-4039(00)00520-7.

[158] Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S., Synthesis of 6-(2-Thienyl)Purine Nucleoside Derivatives That Form Unnatural Base Pairs with Pyridin-2-One Nucleosides. *Bioorganic Med. Chem. Lett.* **2001**, *11*, 2221–2223. https://doi.org/10.1016/S0960-894X(01)00415-2.

[159] Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; et al., An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins. *Nat. Biotechnol.* **2002**, *20*, 177–182. https://doi.org/10.1038/nbt0202-177.

[160] Moriyama, K.; Kimoto, M.; Mitsui, T.; Yokoyama, S.; Hirao, I., Site-Specific Biotinylation of RNA Molecules by Transcription Using Unnatural Base Pairs. *Nucleic Acids Res.* **2005**, *33* (15), e129–e129. https://doi.org/10.1093/nar/gni128.

[161] Kawai, R.; Kimoto, M.; Ikeda, S.; Mitsui, T.; Endo, M.; Yokoyama, S.; Hirao, I., Site-Specific Fluorescent Labeling of RNA Molecules by Specific Transcription Using Unnatural Base Pairs. *J. Am. Chem. Soc.* **2005**, *127*, 17286–17295. https://doi.org/10.1093/nass/49.1.287.

[162] Hirao, I., Placing Extra Components into RNA by Specific Transcription Using Unnatural Base Pair Systems. *Biotechniques* **2006**, *40* (6), 711–717. https://doi.org/10.2144/000112187.

[163] Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S., An Unnatural Hydrophobic Base Pair System: Site-Specific Incorporation of Nucleotide Analogs into DNA and RNA. *Nat. Methods* **2006**, *3* (9), 729–735. https://doi.org/10.1038/nmeth915.

[164] Hirao, I.; Kimoto, M., Unnatural Base Pair Systems toward the Expansion of the Genetic Alphabet in the Central Dogma. *Proc. Japan Acad. Ser. B* **2012**, *88* (7), 345–367. https://doi.org/10.2183/pjab.88.345.

[165] Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I., Highly Specific Unnatural Base Pair Systems as a Third Base Pair for PCR Amplification. *Nucleic Acids Res.* **2012**, *40* (6), 2793–2806. https://doi.org/10.1093/nar/gkr1068.

[166] Kimoto, M.; Sato, A.; Kawai, R.; Yokoyama, S.; Hirao, I., Site-Specific Incorporation of Functional Components into RNA by Transcription Using Unnatural Base Pair Systems. *Nucleic Acids Symp. Ser.* **2009**, *53*, 73–74. https://doi.org/10.1007/978-1-60761-652-8_25.

[167] Leconte, A. M.; Romesberg, F. E., Amplify This! DNA and RNA Get a Third Base Pair. *Nat. Methods* **2006**, *3* (9), 667–668. https://doi.org/10.1038/nmeth0906-667.

[168] Okamoto, I.; Miyatake, Y.; Kimoto, M.; Hirao, I., High Fidelity, Efficiency and Functionalization of Ds-Px Unnatural Base Pairs in PCR Amplification for a Genetic Alphabet Expansion System. *ACS Synth. Biol.* **2016**, *5*, 1220–1230. https://doi.org/10.1021/acssynbio.5b00253.

[169] Hikida, Y.; Kimoto, M.; Hirao, I.; Yokoyama, S., Crystal Structure of Deep Vent DNA Polymerase. *Biochem. Biophys. Res. Commun.* **2017**, *483*, 52–57. https://doi.org/10.1016/j.bbrc.2017.01.007.

[170] Kimoto, M.; Yamashige, R.; Yokoyama, S.; Hirao, I., PCR Amplification and Transcription for Site-Specific Labeling of Large RNA Molecules by a Two-Unnatural-Base-Pair System. *J. Nucleic Acids* **2012**, *2012*, 1–8. https://doi.org/10.1155/2012/230943.

[171] Ishizuka, T.; Kimoto, M.; Sato, A.; Hirao, I., Site-Specific Functionalization of RNA Molecules by an Unnatural Base Pair Transcription System via Click Chemistry. *Chem. Commun.* **2012**, *48* (88), 10835. https://doi.org/10.1039/c2cc36293g.

[172] Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I., Generation of High-Affinity DNA Aptamers Using an Expanded Genetic Alphabet. *Nat. Biotechnol.* **2013**, *31* (5), 453–457. https://doi.org/10.1038/nbt.2556.

[173] Matsunaga, K.; Kimoto, M.; Hanson, C.; Sanford, M.; Young, H. A.; Hirao, I., Architecture of High-Affinity Unnatural-Base DNA Aptamers toward Pharmaceutical Applications. *Sci. Rep.* **2015**, *5*, 18478. https://doi.org/10.1038/srep18478.

[174] Kimoto, M.; Nakamura, M.; Hirao, I., Post-ExSELEX Stabilization of an Unnatural-Base DNA Aptamer Targeting VEGF165 toward Pharmaceutical Applications. *Nucleic Acids Res.* **2016**, *44* (15), 7487–7494. https://doi.org/10.1093/nar/gkw619.

[175] Futami, K.; Kimoto, M.; Lim, Y. W. S.; Hirao, I., Genetic Alphabet Expansion Provides Versatile Specificities and Activities of Unnatural-Base DNA Aptamers Targeting Cancer Cells. *Mol. Ther. - Nucleic Acids* **2019**, *14* (March), 158–170. https://doi.org/10.1016/j.omtn.2018.11.011.

[176] Betz, K.; Kimoto, M.; Diederichs, K.; Hirao, I.; Marx, A., Structural Basis for Expansion of the Genetic Alphabet with an Artificial Nucleobase Pair. *Angew. Chemie Int. Ed.* **2017**, *56* (39), 12000–12003. https://doi.org/10.1002/anie.201704190.

[177] Karalkar, N. B.; Leal, N. A.; Kim, M.-S.; Bradley, K. M.; Benner, S. A., Synthesis and Enzymology of 2′-Deoxy-7-Deazaisoguanosine Triphosphate and Its Complement: A Second Generation Pair in an Artificially Expanded Genetic Information System. *ACS Synth. Biol.* **2016**, *5* (7), 672–678. https://doi.org/10.1021/acssynbio.5b00276.

[178] Berger, M.; Wu, Y.; Ogawa, A. K.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E., Universal Bases for Hybridization, Replication and Chain Termination. *Nucleic Acids Res.* **2000**, *28* (15), 2911–2914. https://doi.org/10.1093/nar/28.15.2911.

[179] Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E., Rational Design of an Unnatural Base Pair with Increased Kinetic Selectivity [14]. *J. Am. Chem. Soc.* **2000**, *122* (36), 8803–8804. https://doi.org/10.1021/ja001450u.

[180] Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E., Efforts toward Expansion of the Genetic Alphabet: Replication of DNA with Three Base Pairs. *J. Am. Chem. Soc.* **2001**, *123* (30), 7439–7440. https://doi.org/10.1021/ja010731e.

[181] Henry, A. A.; Romesberg, F. E., Beyond A, C, G and T: Augmenting Nature's Alphabet. *Curr. Opin. Chem. Biol.* **2003**, *7* (6), 727–733. https://doi.org/10.1016/j.cbpa.2003.10.011.

[182] Henry, A. A.; Yu, C.; Romesberg, F. E., Determinants of Unnatural Nucleobase Stability and Polymerase Recognition. *J. Am. Chem. Soc.* **2003**, *125*, 9638–9646. https://doi.org/10.1021/ja035398o.

[183] Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E., Discovery, Characterization, and Optimization of an Unnatural Base Pair for Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.* **2008**, *130* (7), 2336–2343. https://doi.org/10.1021/ja078223d.

[184] Matsuda, S.; Henry, A. A.; Romesberg, F. E., Optimization of Unnatural Base Pair Packing for Polymerase Recognition. *J. Am. Chem. Soc.* **2006**, *128*, 6369–6375. https://doi.org/10.1021/ja057575m.

[185] Hwang, G. T.; Romesberg, F. E., Unnatural Substrate Repertoire of A, B, and X Family DNA Polymerases. *J. Am. Chem. Soc.* **2008**, *130* (44), 14872–14882. https://doi.org/10.1021/ja803833h.
[186] Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E., Efficient and Sequence-Independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-Letter Genetic Alphabet. *PNAS* **2012**, *109* (30), 12005–12010. https://doi.org/10.1073/pnas.1205176109.
[187] Seo, Y. J.; Matsuda, S.; Romesberg, F. E., Transcription of an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* **2009**, *131* (14), 5046–5047. https://doi.org/10.1021/ja9006996.
[188] Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E., Optimization of an Unnatural Base Pair Toward Natural-Like Replication. *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252. https://doi.org/10.1021/ja807853m.
[189] Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A., KlenTaq Polymerase Replicates Unnatural Base Pairs by Inducing a Watson-Crick Geometry. *Nat. Chem. Biol.* **2012**, *8*, 612–614. https://doi.org/10.1038/nchembio.966.
[190] Malyshev, D. A.; Romesberg, F. E., The Expanded Genetic Alphabet. *Angew. Chemie Int. Ed.* **2015**, *54* (41), 11930–11944. https://doi.org/10.1002/anie.201502890.
[191] Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A., Structural Insights into DNA Replication without Hydrogen Bonds. *J. Am. Chem. Soc.* **2013**, *135* (49), 18637–18643. https://doi.org/10.1021/ja409609j.
[192] Seo, Y. J.; Malyshev, D. A.; Lavergne, T.; Ordoukhanian, P.; Romesberg, F. E., Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* **2011**, *133* (49), 19878–19888. https://doi.org/10.1021/ja207907d.
[193] Dhami, K.; Malyshev, D. A.; Ordoukhanian, P.; Kubelka, T.; Hocek, M.; Romesberg, F. E., Systematic Exploration of a Class of Hydrophobic Unnatural Base Pairs Yields Multiple New Candidates for the Expansion of the Genetic Alphabet. *Nucleic Acids Res.* **2014**, *42* (16), 10235–10244. https://doi.org/10.1093/nar/gku715.
[194] Li, L.; Degardin, M.; Lavergne, T.; Malyshev, D. A.; Dhami, K.; Ordoukhanian, P.; Romesberg, F. E., Natural-Like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications. *J. Am. Chem. Soc.* **2014**, *136*, 826–829. https://doi.org/10.1021/ja408814g.
[195] Morris, S. E.; Feldman, A. W.; Romesberg, F. E., Synthetic Biology Parts for the Storage of Increased Genetic Information in Cells. *ACS Synth. Biol.* **2017**, *6* (10), 1834–1840. https://doi.org/10.1021/acssynbio.7b00115.
[196] Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa Jr., I. R.; Romesberg, F. E., A Semi-Synthetic Organism with an Expanded Genetic Alphabet. *Nature* **2014**, *509*, 385–388. https://doi.org/10.1038/nature13314.
[197] Romesberg, F. E.; Malyshev, D. A. Import of Unnatural or Modified Nucleoside Triphosphates into Cells via Nucleic Acid Triphosphate Transporters, 2017.
[198] Feldman, A. W.; Fischer, E. C.; Ledbetter, M. P.; Liao, J.-Y.; Chaput, J. C.; Romesberg, F. E., A Tool for the Import of Natural and Unnatural Nucleoside Triphosphates into Bacteria. *J. Am. Chem. Soc.* **2018**, *140* (4), 1447–1454. https://doi.org/10.1021/jacs.7b11404.
[199] Zhang, Y.; Lamb, B. M.; Feldman, A. W.; Zhou, A. X.; Lavergne, T.; Li, L.; Romesberg, F. E., A Semisynthetic Organism Engineered for the Stable Expansion of the Genetic Alphabet. *Proc. Natl. Acad. Sci.* **2017**, *114* (6), 1317–1322. https://doi.org/10.1073/pnas.1616443114.
[200] Ledbetter, M. P.; Karadeema, R. J.; Romesberg, F. E., Reprograming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.* **2018**, *140* (2), 758–765. https://doi.org/10.1021/jacs.7b11488.

[201] Zhang, Y.; Ptacin, J. L.; Fischer, E. C.; Aerni, H. R.; Caffaro, C. E.; San Jose, K.; Feldman, A. W.; Turner, C. R.; Romesberg, F. E., A Semi-Synthetic Organism That Stores and Retrieves Increased Genetic Information. *Nature* **2017**, *551*, 644–647. https://doi.org/10.1038/nature24659.

[202] Dien, V. T.; Morris, S. E.; Karadeema, R. J.; Romesberg, F. E., Expansion of the Genetic Code via Expansion of the Genetic Alphabet. *Curr. Opin. Chem. Biol.* **2018**, *46*, 196–202. https://doi.org/10.1016/j.cbpa.2018.08.009.

[203] Zhang, Y.; Romesberg, F. E., Semisynthetic Organisms with Expanded Genetic Codes. *Biochemistry* **2018**, *57* (15), 2177–2178. https://doi.org/10.1021/acs.biochem.8b00013.

[204] Feldman, A. W.; Dien, V. T.; Karadeema, R. J.; Fischer, E. C.; You, Y.; Anderson, B. A.; Krishnamurthy, R.; Chen, J. S.; Li, L.; Romesberg, F. E., Optimization of Replication, Transcription, and Translation in a Semi-Synthetic Organism. *J. Am. Chem. Soc.* **2019**, *141* (27), 10644–10653. https://doi.org/10.1021/jacs.9b02075.

[205] Lavergne, T.; Lamichhane, R.; Malyshev, D. A.; Li, Z.; Li, L.; Sperling, E.; Williamson, J. R.; Millar, D. P.; Romesberg, F. E., FRET Characterization of Complex Conformational Changes in a Large 16S Ribosomal RNA Fragment Site-Specifically Labeled Using Unnatural Base Pairs. *ACS Chem. Biol.* **2016**, *11*, 1347–1353. https://doi.org/10.1021/acschembio.5b00952.

[206] Knall, A.-C.; Kovačič, S.; Hollauf, M.; Reishofer, D.; Saf, R.; Slugovc, C., Inverse Electron Demand Diels–Alder (IEDDA) Functionalisation of Macroporous Poly(Dicyclopentadiene) Foams. *Chem. Commun.* **2013**, *49* (66), 7325. https://doi.org/10.1039/c3cc42925c.

[207] Knall, A.-C.; Slugovc, C., Inverse Electron Demand Diels–Alder (IEDDA)-Initiated Conjugation: A (High) Potential Click Chemistry Scheme. *Chem. Soc. Rev.* **2013**, *42* (12), 5131. https://doi.org/10.1039/c3cs60049a.

[208] Knall, A.-C.; Hollauf, M.; Slugovc, C., Kinetic Studies of Inverse Electron Demand Diels–Alder Reactions (IEDDA) of Norbornenes and 3,6-Dipyridin-2-Yl-1,2,4,5-Tetrazine. *Tetrahedron Lett.* **2014**, *55* (34), 4763–4766. https://doi.org/10.1016/j.tetlet.2014.07.002.

[209] Mayer, S.; Lang, K., Tetrazines in Inverse-Electron-Demand Diels-Alder Cycloadditions and Their Use in Biology. *Synthesis (Stuttg).* **2017**, *49*, 830–848. https://doi.org/10.1055/s-0036-1588682.

[210] Kozma, E.; Demeter, O.; Kele, P., Bio-Orthogonal Fluorescent Labelling of Biopolymers through Inverse-Electron-Demand Diels–Alder Reactions. *ChemBioChem* **2017**, *18*, 486–501. https://doi.org/10.1002/cbic.201700096.

[211] Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **2001**, *40*, 2004–2021. https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

[212] Sletten, E. M.; Bertozzi, C. R., From Mechanism to Mouse: A Tale of Two Bioorthogonal Reactions. *Acc. Chem. Res.* **2011**, *44* (9), 666–676. https://doi.org/10.1021/ar200148z.

[213] Hermann, T.; Heumann, H., Structure and Distance Determination in RNA with Copper Phenanthroline Probing; 2000; Vol. 318, pp 33–43. https://doi.org/10.1016/S0076-6879(00)18042-5.

[214] Paredes, E.; Das, S. R., Click Chemistry for Rapid Labeling and Ligation of RNA. *ChemBioChem* **2011**, *12* (1), 125–131. https://doi.org/10.1002/cbic.201000466.

[215] Gong, Y. H.; Miomandre, F.; Méallet-Renault, R.; Badré, S.; Galmiche, L.; Tang, J.; Audebert, P.; Clavier, G., Synthesis and Physical Chemistry of S-Tetrazines: Which Ones Are Fluorescent and Why? *European J. Org. Chem.* **2009**, *2009* (35), 6121–6128. https://doi.org/10.1002/ejoc.200900964.

[216] Devaraj, N. K.; Weissleder, R., Biomedical Applications of Tetrazine Cycloadditions. *Acc. Chem. Res.* **2011**, *44* (9), 816–827. https://doi.org/10.1021/ar200037t.

[217] Wieczorek, A.; Buckup, T.; Wombacher, R., Rigid Tetrazine Fluorophore Conjugates with Fluorogenic Properties in the Inverse Electron Demand Diels–Alder Reaction. *Org. Biomol. Chem.* **2014**, *12*, 4177–4185. https://doi.org/10.1039/C4OB00245H.

[218] Meimetis, L. G.; Carlson, J. C. T.; Giedt, R. J.; Kohler, R. H.; Weissleder, R., Ultrafluorogenic Coumarin-Tetrazine Probes for Real-Time Biological Imaging. *Angew. Chemie Int. Ed.* **2014**, *53* (29), 7531–7534. https://doi.org/10.1002/anie.201403890.

[219] Devaraj, N. K.; Hilderbrand, S.; Upadhyay, R.; Mazitschek, R.; Weissleder, R., Bioorthogonal Turn-On Probes for Imaging Small Molecules Inside Living Cells. *Angew. Chemie* **2010**, *122*, 2931–2934. https://doi.org/10.1002/anie.200906120.

[220] Siegl, S. J.; Galeta, J.; Dzijak, R.; Vázquez, A.; Del Río-Villanueva, M.; Dračínský, M.; Vrabel, M., An Extended Approach for the Development of Fluorogenic Trans -Cyclooctene–Tetrazine Cycloadditions. *ChemBioChem* **2019**, *20* (7), 886–890. https://doi.org/10.1002/cbic.201800711.

[221] Carlson, J. C. T.; Meimetis, L. G.; Hilderbrand, S. A.; Weissleder, R., BODIPY-Tetrazine Derivatives as Superbright Bioorthogonal Turn-on Probes. *Angew. Chemie* **2013**, *125*, 7055–7058. https://doi.org/10.1002/anie.201301100.

[222] Devaraj, N. K.; Upadhyay, R.; Haun, J. B.; Hilderbrand, S. A.; Weissleder, R., Fast and Sensitive Pretargeted Labeling of Cancer Cells through a Tetrazine/Trans-Cyclooctene Cycloaddition. *Angew. Chemie - Int. Ed.* **2009**, *48*, 7013–7016. https://doi.org/10.1002/anie.200903233.

[223] Taylor, M. T.; Blackman, M. L.; Dmitrenko, O.; Fox, J. M., Design and Synthesis of Highly Reactive Dienophiles for the Tetrazine-Trans-Cyclooctene Ligation. *J. Am. Chem. Soc.* **2011**, *133*, 9646–9649. https://doi.org/10.1021/ja201844c.

[224] Blizzard, R. J.; Backus, D. R.; Brown, W.; Bazewicz, C. G.; Li, Y.; Mehl, R. A., Ideal Bioorthogonal Reactions Using A Site-Specifically Encoded Tetrazine Amino Acid. *J. Am. Chem. Soc.* **2015**, *137*, 10044–10047. https://doi.org/10.1021/jacs.5b03275.

[225] Machida, T.; Lang, K.; Xue, L.; Chin, J. W.; Winssinger, N., Site-Specific Glycoconjugation of Protein via Bioorthogonal Tetrazine Cycloaddition with a Genetically Encoded Trans -Cyclooctene or Bicyclononyne. *Bioconjug. Chem.* **2015**, *26*, 802–806. https://doi.org/10.1021/acs.bioconjchem.5b00101.

[226] Asare-Okai, P. N.; Agustin, E.; Fabris, D.; Royzen, M., Site-Specific Fluorescence Labelling of RNA Using Bio-Orthogonal Reaction of Trans-Cyclooctene and Tetrazine. *Chem. Commun.* **2014**, *50* (58), 7844. https://doi.org/10.1039/c4cc02435d.

[227] Rossin, R.; van den Bosch, S. M.; ten Hoeve, W.; Carvelli, M.; Versteegen, R. M.; Lub, J.; Robillard, M. S., Highly Reactive Trans-Cyclooctene Tags with Improved Stability for Diels-Alder Chemistry in Living Systems. *Bioconjug. Chem.* **2013**, *24*, 1210–1217. https://doi.org/10.1021/bc400153y.

[228] Selvaraj, R.; Fox, J. M., Trans-Cyclooctene - a Stable, Voracious Dienophile for Bioorthogonal Labeling. *Curr. Opin. Chem. Biol.* **2013**, *17*, 753–760. https://doi.org/10.1016/j.cbpa.2013.07.031.

[229] Royzen, M.; Yap, G. P. A.; Fox, J. M., A Photochemical Synthesis of Functionalized Trans-Cyclooctenes Driven by Metal Complexation. *J. Am. Chem. Soc.* **2008**, *130*, 3760–3761. https://doi.org/10.1021/ja8001919.

[230] Svatunek, D.; Denk, C.; Rosecker, V.; Sohr, B.; Hametner, C.; Allmaier, G.; Fröhlich, J.; Mikula, H., Efficient Low-Cost Preparation of Trans-Cyclooctenes Using a Simplified Flow Setup for Photoisomerization. *Monatshefte fur Chemie* **2016**, *147*, 579–585. https://doi.org/10.1007/s00706-016-1668-z.

[231] Schoch, J.; Wiessler, M.; Jäschke, A., Post-Synthetic Modification of DNA by Inverse-Electron-Demand Diels-Alder Reaction. *J. Am. Chem. Soc.* **2010**, *132*, 8846–8847. https://doi.org/10.1021/ja102871p.

[232] Muttach, F.; Muthmann, N.; Reichert, D.; Anhäuser, L.; Rentmeister, A,. A Benzylic Linker Promotes Methyltransferase Catalyzed Norbornene Transfer for Rapid Bioorthogonal Tetrazine Ligation. *Chem. Sci.* **2017**, *8*, 7947–7953. https://doi.org/10.1039/c7sc03631k.

[233] Plass, T.; Milles, S.; Koehler, C.; Szymański, J.; Mueller, R.; Wießler, M.; Schultz, C.; Lemke, E. A., Amino Acids for Diels-Alder Reactions in Living Cells. *Angew. Chemie - Int. Ed.* **2012**, *51*, 4166–4170. https://doi.org/10.1002/anie.201108231.

[234] Pyka, A. M.; Domnick, C.; Braun, F.; Kath-Schorr, S., Diels–Alder Cycloadditions on Synthetic RNA in Mammalian Cells. *Bioconjug. Chem.* **2014**, *25* (8), 1438–1443. https://doi.org/10.1021/bc500302y.

[235] Domnick, C.; Eggert, F.; Kath-Schorr, S., Site-Specific Enzymatic Introduction of a Norbornene Modified Unnatural Base into RNA and Application in Post-Transcriptional Labeling. *Chem. Commun.* **2015**, *51*, 8253–8256. https://doi.org/10.1039/C5CC01765C.

[236] Engelsma, S. B.; Willems, L. I.; van Paaschen, C. E.; van Kasteren, S. I.; van der Marel, G. A.; Overkleeft, H. S.; Filippov, D. V., Acylazetine as a Dienophile in Bioorthogonal Inverse Electron-Demand Diels-Alder Ligation. *Org. Lett.* **2014**, *16*, 2744–2747. https://doi.org/10.1021/ol501049c.

[237] Li, Z.; Wang, D.; Li, L.; Pan, S.; Na, Z.; Tan, C. Y. J.; Yao, S. Q., "Minimalist" Cyclopropene-Containing Photo-Cross-Linkers Suitable for Live-Cell Imaging and Affinity-Based Protein Labeling. *J. Am. Chem. Soc.* **2014**, *136*, 9990–9998. https://doi.org/10.1021/ja502780z.

[238] Kamber, D. N.; Nazarova, L. A.; Liang, Y.; Lopez, S. A.; Patterson, D. M.; Shih, H. W.; Houk, K. N.; Prescher, J. A., Isomeric Cyclopropenes Exhibit Unique Bioorthogonal Reactivities. *J. Am. Chem. Soc.* **2013**, *135*, 13680–13683. https://doi.org/10.1021/ja407737d.

[239] Yang, J.; Šečkute, J.; Cole, C. M.; Devaraj, N. K., Live-Cell Imaging of Cyclopropene Tags with Fluorogenic Tetrazine Cycloadditions. *Angew. Chemie* **2012**, *124*, 7594–7597. https://doi.org/10.1002/anie.201202122.

[240] Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A., Functionalized Cyclopropenes As Bioorthogonal Chemical Reporters. *J. Am. Chem. Soc.* **2012**, *134*, 18638–18643. https://doi.org/10.1021/ja3060436.

[241] Späte, A.-K.; Bußkamp, H.; Niederwieser, A.; Schart, V. F.; Marx, A.; Wittmann, V., Rapid Labeling of Metabolically Engineered Cell-Surface Glycoconjugates with a Carbamate-Linked Cyclopropene Reporter. *Bioconjug. Chem.* **2014**, *25*, 147–154. https://doi.org/10.1021/bc4004487.

[242] Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q., Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry-Mediated Protein Labeling in Mammalian Cells. *Angew. Chemie* **2012**, *124*, 10752–10756. https://doi.org/10.1002/anie.201205352.

[243] Ramil, C. P.; Dong, M.; An, P.; Lewandowski, T. M.; Yu, Z.; Miller, L. J.; Lin, Q., Spirohexene-Tetrazine Ligation Enables Bioorthogonal Labeling of Class B G Protein-Coupled Receptors in Live Cells. *J. Am. Chem. Soc.* **2017**, *139*, 13376–13386. https://doi.org/10.1021/jacs.7b05674.

[244] Protopopova, M. N.; Shapiro, E. A., Carbene Synthesis and Chemical Reactions of Cyclopropene-3-Carboxylate Esters — Promising Intermediates for Organic Synthesis. *Russ. Chem. Rev.* **1989**, *58* (7), 667–681. https://doi.org/10.1070/RC1989v058n07ABEH003469.

[245] Yan, N.; Liu, X.; Pallerla, M. K.; Fox, J. M., Synthesis of Stable Derivatives of Cycloprop-2-Ene Carboxylic Acid. *J. Org. Chem.* **2008**, *73*, 4283–4286. https://doi.org/10.1021/jo800042w.

[246] Morandi, B.; Carreira, E. M., Rhodium-Catalyzed Cyclopropenation of Alkynes: Synthesis of Trifluoromethyl-Substituted Cyclopropenes. *Angew. Chemie* **2010**, *122*, 4390–4392. https://doi.org/10.1002/anie.201000787.

[247] Vicente, R., Recent Progresses towards the Strengthening of Cyclopropene Chemistry. *Synthesis (Stuttg).* **2016**, *48* (15), 2343–2360. https://doi.org/10.1055/s-0035-1561644.

[248] Horton, T. E.; Clardy, D. R.; DeRose, V. J., Electron Paramagnetic Resonance Spectroscopic Measurement of Mn2+ Binding Affinities to the Hammerhead Ribozyme and Correlation with Cleavage Activity. *Biochemistry* **1998**, *37*, 18094–18101. https://doi.org/10.1021/bi981425p.

[249] Morrissey, S. R.; Horton, T. E.; DeRose, V. J., Mn2+ Sites in the Hammerhead Ribozyme Investigated by EPR and Continuous-Wave Q-Band ENDOR Spectroscopies. *J. Am. Chem. Soc.* **2000**, *122*, 3473–3481. https://doi.org/10.1021/ja992989z.
[250] Prisner, T.; Rohrer, M.; MacMillan, F., Pulsed EPR Spectroscopy: Biological Applications. *Annu. Rev. Phys. Chem.* **2001**, *52* (1), 279–313. https://doi.org/10.1146/annurev.physchem.52.1.279.
[251] Klare, J. P.; Steinhoff, H.-J., Spin Labeling EPR. *Photosynth. Res.* **2009**, *102* (2–3), 377–390. https://doi.org/10.1007/s11120-009-9490-7.
[252] Duss, O.; Yulikov, M.; Jeschke, G.; Allain, F. H.-T., EPR-Aided Approach for Solution Structure Determination of Large RNAs or Protein-RNA Complexes. *Nat. Commun.* **2014**, *5*, 1–9. https://doi.org/10.1038/ncomms4669.
[253] Haugland, M. M.; Lovett, J. E.; Anderson, E. A., Advances in the Synthesis of Nitroxide Radicals for Use in Biomolecule Spin Labelling. *Chem. Soc. Rev.* **2018**, *47*, 668–680. https://doi.org/10.1039/C6CS00550K.
[254] Jahromy, Y. N.; Schubert, E., Demystifying EPR: A Rookie Guide to the Application of Electron Paramagnetic Resonance Spectroscopy on Biomolecules. *Prog. Biol. Sci.* **2014**, *4* (2), 133–152.
[255] Edwards, D. T.; Ma, Z.; Meade, T. J.; Goldfarb, D.; Han, S.; Sherwin, M. S., Extending the Distance Range Accessed with Continuous Wave EPR with Gd3+ Spin Probes at High Magnetic Fields. *Phys. Chem. Chem. Phys.* **2013**, *15* (27), 11313. https://doi.org/10.1039/c3cp43787f.
[256] Sale, K.; Song, L.; Liu, Y. S.; Perozo, E.; Fajer, P., Explicit Treatment of Spin Labels in Modeling of Distance Constraints from Dipolar EPR and DEER. *J. Am. Chem. Soc.* **2005**, *127* (26), 9334–9335. https://doi.org/10.1021/ja051652w.
[257] Jeschke, G., Distance Measurements in the Nanometer Range by Pulse EPR. *ChemPhysChem* **2002**, *3* (11), 927–932.
[258] Schiemann, O.; Piton, N.; Plackmeyer, J.; Bode, B. E.; Prisner, T. F.; Engels, J. W., Spin Labeling of Oligonucleotides with the Nitroxide TPA and Use of PELDOR, a Pulse EPR Method, to Measure Intramolecular Distances. *Nat. Protoc.* **2007**, *2* (4), 904–923. https://doi.org/10.1038/nprot.2007.97.
[259] Jassoy, J. J.; Berndhaeuser, A.; Duthie, F.; Kuehn, S. P.; Hagelueken, G.; Schiemann, O., Versatile Trityl Spin Labels for Nanometer Distance Measurements on Biomolecules In Vitro and within Cells. *Angew. Chemie - Int. Ed.* **2017**, *56*, 177–181. https://doi.org/10.1002/anie.201609085.
[260] Yin, D. M.; Hammler, D.; Peter, M. F.; Marx, A.; Schmitz, A.; Hagelueken, G., Inhibitor-Directed Spin Labelling-A High Precision and Minimally Invasive Technique to Study the Conformation of Proteins in Solution. *Chem. - A Eur. J.* **2018**, *24* (26), 6665–6671. https://doi.org/10.1002/chem.201706047.
[261] Schiemann, O., Intracellular Ruler - EPR-Based Structural Elucidation of Biomolecules. *Lab. J.* **2018**.
[262] Büttner, L.; Seikowski, J.; Wawrzyniak, K.; Ochmann, A.; Höbartner, C., Synthesis of Spin-Labeled Riboswitch RNAs Using Convertible Nucleosides and DNA-Catalyzed RNA Ligation. *Bioorganic Med. Chem.* **2013**, *21*, 6171–6180. https://doi.org/10.1016/j.bmc.2013.04.007.
[263] Karthikeyan, G.; Bonucci, A.; Casano, G.; Gerbaud, G.; Abel, S.; Thomé, V.; Kodjabachian, L.; Magalon, A.; Guigliarelli, B.; Belle, V.; et al., A Bioresistant Nitroxide Spin Label for In-Cell EPR Spectroscopy: In Vitro and In Oocytes Protein Structural Dynamics Studies. *Angew. Chemie Int. Ed.* **2018**, *57* (5), 1366–1370. https://doi.org/10.1002/anie.201710184.
[264] Gophane, D. B.; Endeward, B.; Prisner, T. F.; Sigurdsson, S. T., A Semi-Rigid Isoindoline-Derived Nitroxide Spin Label for RNA. *Org. Biomol. Chem.* **2018**, *16*, 816–824. https://doi.org/10.1039/C7OB02870A.

[265] Kerzhner, M.; Matsuoka, H.; Wuebben, C.; Famulok, M.; Schiemann, O., High-Yield Spin Labeling of Long RNAs for Electron Paramagnetic Resonance Spectroscopy. *Biochemistry* **2018**, *57* (20), 2923–2931. https://doi.org/10.1021/acs.biochem.8b00040.

[266] Saha, S.; Jagtap, A. P.; Sigurdsson, S. T., Site-Directed Spin Labeling of 2′-Amino Groups in RNA with Isoindoline Nitroxides That Are Resistant to Reduction. *Chem. Commun.* **2015**, *51* (66), 13142–13145. https://doi.org/10.1039/C5CC05014F.

[267] Schiemann, O.; Weber, A.; Edwards, T. E.; Prisner, T. F.; Sigurdsson, S. T., Nanometer Distance Measurements on RNA Using PELDOR. *J. Am. Chem. Soc.* **2003**, *125*, 3434–3435. https://doi.org/10.1021/ja0274610.

[268] Kaminker, I.; Bye, M.; Mendelman, N.; Gislason, K.; Sigurdsson, S. T.; Goldfarb, D., Distance Measurements between Manganese(II) and Nitroxide Spin-Labels by DEER Determine a Binding Site of Mn2+ in the HP92 Loop of Ribosomal RNA. *Phys. Chem. Chem. Phys.* **2015**, *17*, 15098–15102. https://doi.org/10.1039/C5CP01624J.

[269] Halbmair, K.; Seikowski, J.; Tkach, I.; Höbartner, C.; Sezer, D.; Bennati, M., High-Resolution Measurement of Long-Range Distances in RNA: Pulse EPR Spectroscopy with TEMPO-Labeled Nucleotides. *Chem. Sci.* **2016**, *7*, 3172–3180. https://doi.org/10.1039/C5SC04631A.

[270] Krstić, I.; Hänsel, R.; Romainczyk, O.; Engels, J. W.; Dötsch, V.; Prisner, T. F., Long-Range Distance Measurements on Nucleic Acids in Cells by Pulsed EPR Spectroscopy. *Angew. Chemie - Int. Ed.* **2011**, *50*, 5070–5074. https://doi.org/10.1002/anie.201100886.

[271] Grytz, C. M.; Marko, A.; Cekan, P.; Sigurdsson, S. T.; Prisner, T. F., Flexibility and Conformation of the Cocaine Aptamer Studied by PELDOR. *Phys. Chem. Chem. Phys.* **2016**, *18* (4), 2993–3002. https://doi.org/10.1039/C5CP06158J.

[272] Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantirek, L.; Dötsch, V., In-Cell NMR and EPR Spectroscopy of Biomacromolecules. *Angew. Chemie Int. Ed.* **2014**, *53*, 10300–10314. https://doi.org/10.1002/anie.201311320.

[273] Azarkh, M.; Okle, O.; Singh, V.; Seemann, I. T.; Hartig, J. S.; Dietrich, D. R.; Drescher, M., Long-Range Distance Determination in a DNA Model System inside Xenopus Laevis Oocytes by In-Cell Spin-Label EPR. *ChemBioChem* **2011**, *12* (13), 1992–1995. https://doi.org/10.1002/cbic.201100281.

[274] Babaylova, E. S.; Ivanov, A. V.; Malygin, A. A.; Vorobjeva, M. A.; Venyaminova, A. G.; Polienko, Y. F.; Kirilyuk, I. A.; Krumkacheva, O. A.; Fedin, M. V.; Karpova, G. G.; et al., A Versatile Approach for Site-Directed Spin Labeling and Structural EPR Studies of RNAs. *Org. Biomol. Chem.* **2014**, *12* (19), 3129. https://doi.org/10.1039/c3ob42154f.

[275] Babaylova, E. S.; Malygin, A. A.; Lomzov, A. A.; Pyshnyi, D. V; Yulikov, M.; Jeschke, G.; Krumkacheva, O. A.; Fedin, M. V; Karpova, G. G.; Bagryanskaya, E. G., Complementary-Addressed Site-Directed Spin Labeling of Long Natural RNAs. *Nucleic Acids Res.* **2016**, *44* (16), 7935–7943. https://doi.org/10.1093/nar/gkw516.

[276] Sauer, M., Reversible Molecular Photoswitches: A Key Technology for Nanoscience and Fluorescence Imaging. *Proc. Natl. Acad. Sci.* **2005**, *102* (27), 9433–9434. https://doi.org/10.1073/pnas.0504264102.

[277] Russew, M.-M.; Hecht, S., Photoswitches: From Molecules to Materials. *Adv. Mater.* **2010**, *22* (31), 3348–3360. https://doi.org/10.1002/adma.200904102.

[278] Babii, O.; Afonin, S.; Berditsch, M.; Reißer, S.; Mykhailiuk, P. K.; Kubyshkin, V. S.; Steinbrecher, T.; Ulrich, A. S.; Komarov, I. V., Controlling Biological Activity with Light: Diarylethene-Containing Cyclic Peptidomimetics. *Angew. Chemie Int. Ed.* **2014**, *53* (13), 3392–3395. https://doi.org/10.1002/anie.201310019.

[279] Buckup, T.; Sarter, C.; Volpp, H.-R.; Jäschke, A.; Motzkus, M., Ultrafast Time-Resolved Spectroscopy of Diarylethene-Based Photoswitchable Deoxyuridine Nucleosides. *J. Phys. Chem. Lett.* **2015**, *6* (23), 4717–4721. https://doi.org/10.1021/acs.jpclett.5b01949.

[280] Moreno, J.; Schweighöfer, F.; Wachtveitl, J.; Hecht, S., Reversible Photomodulation of Electronic Communication in a Pi-Conjugated Photoswitch-Fluorophore Molecular Dyad. *Chem. - A Eur. J.* **2016**, *22*, 1070–1075. https://doi.org/10.1002/chem.201503419.

[281] Pu, S.-Z.; Sun, Q.; Fan, C.-B.; Wang, R.-J.; Liu, G., Recent Advances in Diarylethene-Based Multi-Responsive Molecular Switches. *J. Mater. Chem. C* **2016**, *4*, 3075–3093. https://doi.org/10.1039/C6TC00110F.

[282] Merino, E., Synthesis of Azobenzenes: The Coloured Pieces of Molecular Materials. *Chem. Soc. Rev.* **2011**, *40* (7), 3835. https://doi.org/10.1039/c0cs00183j.

[283] Sadovski, O.; Beharry, A. A.; Zhang, F.; Woolley, G. A., Spectral Tuning of Azobenzene Photoswitches for Biological Applications. *Angew. Chemie Int. Ed.* **2009**, *48* (8), 1484–1486. https://doi.org/10.1002/anie.200805013.

[284] Beharry, A. A.; Woolley, G. A., Azobenzene Photoswitches for Biomolecules. *Chem. Soc. Rev.* **2011**, *40* (8), 4422. https://doi.org/10.1039/c1cs15023e.

[285] Szymański, W.; Beierle, J. M.; Kistemaker, H. A. V.; Velema, W. A.; Feringa, B. L., Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chem. Rev.* **2013**, *113* (8), 6114–6178. https://doi.org/10.1021/cr300179f.

[286] Adam, V.; Prusty, D. K.; Centola, M.; Škugor, M.; Hannam, J. S.; Valero, J.; Klöckner, B.; Famulok, M., Expanding the Toolbox of Photoswitches for DNA Nanotechnology Using Arylazopyrazoles. *Chem. - A Eur. J.* **2018**, *24* (5), 1062–1066. https://doi.org/10.1002/chem.201705500.

[287] Haydell, M. W.; Centola, M.; Adam, V.; Valero, J.; Famulok, M., Temporal and Reversible Control of a DNAzyme by Orthogonal Photoswitching. *J. Am. Chem. Soc.* **2018**, *140* (49), 16868–16872. https://doi.org/10.1021/jacs.8b08738.

[288] Prusty, D. K.; Adam, V.; Zadegan, R. M.; Irsen, S.; Famulok, M., Supramolecular Aptamer Nano-Constructs for Receptor-Mediated Targeting and Light-Triggered Release of Chemotherapeutics into Cancer Cells. *Nat. Commun.* **2018**, *9* (1), 535. https://doi.org/10.1038/s41467-018-02929-2.

[289] Škugor, M.; Valero, J.; Murayama, K.; Centola, M.; Asanuma, H.; Famulok, M., Orthogonally Photocontrolled Non-Autonomous DNA Walker. *Angew. Chemie Int. Ed.* **2019**, *58* (21), 6948–6951. https://doi.org/10.1002/anie.201901272.

[290] Bochmann, S.; Mörtel, M.; Witt, A.; Khusniyarov, M. M.; Bachmann, J.; Heinemann, F. W., Synthesis, Characterization, and Properties of Iron(II) Spin-Crossover Molecular Photoswitches Functioning at Room Temperature. *Inorg. Chem.* **2017**, *56* (21), 13174–13186. https://doi.org/10.1021/acs.inorgchem.7b01952.

[291] Chen, H.; Cheng, N.; Ma, W.; Li, M.; Hu, S.; Gu, L.; Meng, S.; Guo, X., Design of a Photoactive Hybrid Bilayer Dielectric for Flexible Nonvolatile Organic Memory Transistors. *ACS Nano* **2016**, *10*, 436–445. https://doi.org/10.1021/acsnano.5b05313.

[292] Jung, H. Y.; You, S.; Lee, C.; You, S.; Kim, Y., One-Pot Synthesis of Monodispersed Silica Nanoparticles for Diarylethene-Based Reversible Fluorescence Photoswitching in Living Cells. *Chem. Commun.* **2013**, *49* (68), 7528–7530. https://doi.org/10.1039/c3cc42852d.

[293] Roubinet, B.; Bossi, M. L.; Alt, P.; Leutenegger, M.; Shojaei, H.; Schnorrenberg, S.; Nizamov, S.; Irie, M.; Belov, V. N.; Hell, S. W., Carboxylierte Photoschaltbare Diarylethene Als Biomarkierungen Für Hochauflösende RESOLFT-Mikroskopie. *Angew. Chemie* **2016**, *128* (49), 15670–15670. https://doi.org/10.1002/ange.201610439.

[294] Wilson, D.; Li, J. W.; Branda, N. R., Visible-Light-Triggered Activation of a Protein Kinase Inhibitor. *ChemMedChem* **2017**, *12* (4), 284–287. https://doi.org/10.1002/cmdc.201600632.

[295] Singer, M.; Jäschke, A., Reversibly Photoswitchable Nucleosides: Synthesis and Photochromic Properties of Diarylethene-Functionalized 7-Deazaadenosine Derivatives. *J. Am. Chem. Soc.* **2010**, *132*, 8372–8377. https://doi.org/10.1021/ja1024782.

[296] Cahová, H.; Jäschke, A., Nucleoside-Based Diarylethene Photoswitches and Their Facile Incorporation into Photoswitchable DNA. *Angew. Chemie - Int. Ed.* **2013**, *52*, 3186–3190. https://doi.org/10.1002/anie.201209943.

[297] Yang, J.; Liang, Y.; Šečkutė, J.; Houk, K. N.; Devaraj, N. K., Synthesis and Reactivity Comparisons of 1-Methyl-3-Substituted Cyclopropene Mini-Tags for Tetrazine Bioorthogonal Reactions. *Chem. - A Eur. J.* **2014**, *20*, 3365–3375. https://doi.org/10.1002/chem.201304225.

[298] Eggert, F., Chemical Modification of an Unnatural Base Pair for the Site-Specific Introduction of Strained Alkene Moieties into RNA, Master Thesis, University of Bonn 2014.

[299] Petiniot, N.; Anciaux, A. J.; Noels, A. F.; Hubert, A. J.; Teyssié, P., Rhodium Catalysed Cyclopropenation of Acetylenes. *Tetrahedron Lett.* **1978**, *14*, 1239–1242. https://doi.org/10.1016/S0040-4039(01)94511-3.

[300] Liao, L.; Zhang, F.; Yan, N.; Golen, J. A.; Fox, J. M., An Efficient and General Method for Resolving Cyclopropene Carboxylic Acids. *Tetrahedron* **2004**, *60*, 1803–1816. https://doi.org/10.1016/j.tet.2003.12.042.

[301] Espino, C. G.; Fiori, K. W.; Kim, M.; Du Bois, J., Expanding the Scope of C−H Amination through Catalyst Design. *J. Am. Chem. Soc.* **2004**, *126*, 15378–15379. https://doi.org/10.1021/ja0446294.

[302] Carter, F. L.; Frampton, V. L,. Review of the Chemistry of Cyclopropene Compounds. *Chem. Rev.* **1964**, *64* (5), 497–525. https://doi.org/10.1021/cr60231a001.

[303] Brückner, R., *Reaktionsmechanismen*, 3rd ed., Spektrum: Heidelberg, **2004**, p. 788.

[304] Fieser, L. F.; Fieser, M., *Reagents for Organic Synthesis*, Vol.1; Wiley: New York, 1967, p. 983.

[305] Merkel, M.; Arndt, S.; Ploschik, D.; Cserép, G. B.; Wenge, U.; Kele, P.; Wagenknecht, H.-A., Scope and Limitations of Typical Copper-Free Bioorthogonal Reactions with DNA: Reactive 2'-Deoxyuridine Triphosphates for Postsynthetic Labeling. *J. Org. Chem.* **2016**, *81*, 7527–7538. https://doi.org/10.1021/acs.joc.6b01205.

[306] New, J. S.; Christopher, W. L.; Yevich, J. P.; Butler, R.; Schlemmer Jr., R. F.; VanderMaelen, C. P.; Cipollina, J. A., The Thieno[3,2-c]Pyridine and Furo[3,2-c]Pyridine Rings: New Pharmacophores with Potential Antipsychotic Activity. *J. Med. Chem.* **1989**, *32* (6), 1147–1156. https://doi.org/10.1017/CBO9781107415324.004.

[307] Ewing, W. R.; Becker, M. R.; Choi-Sledeski, Y. M.; Pauls, H. W.; McGarry, D. G.; Davis, R. S.; Spada, A. P., Substituted Sulfonic Acid N-[(Aminoiminomethyl) Phenylalkyl]-Azaheterocyclamide Compounds, 1998.

[308] Pokhodylo, N. T.; Shyyka, O. Y.; Obushak, N. D., Ethyl 2-Aminothiophene-3-Carboxylates in the Synthesis of Isomeric Thienopyridines. *Chem. Heterocycl. Compd.* **2015**, *50* (12), 1748–1755. https://doi.org/10.1007/s10593-015-1647-z.

[309] Vovk, M. V.; Mel'nichenko, N. V.; Chornous, V. A.; Bratenko, M. K., Intramolecular Cyclization of 4-Isocyanato-3-(2-Naphthyl)-1-Phenylpyrazole Under Friedel-Crafts Reaction Conditions. *Chem. Heterocycl. Compd.* **2002**, *38* (9), 1096–1097.

[310] Zhang, P.; Hu, H.-R.; Huang, Z.-H.; Lei, J.-Y.; Chu, Y.; Ye, D.-Y., Identification of Novel Scaffold of Benzothiazepinones as Non-ATP Competitive Glycogen Synthase Kinase-3b Inhibitors through Virtual Screening. *Bioorganic Med. Chem. Lett.* **2012**, *22*, 7232–7236. https://doi.org/10.1016/j.bmcl.2012.09.043.

[311] Zhang, P.; Hu, H.-R.; Bian, S.-H.; Huang, Z.-H.; Chu, Y.; Ye, D.-Y., Design, Synthesis and Biological Evaluation of Benzothiazepinones (BTZs) as Novel Non-ATP Competitive Inhibitors of Glycogen Synthase Kinase-3b (GSK-3b). *Eur. J. Med. Chem.* **2013**, *61*, 95–103. https://doi.org/10.1016/j.ejmech.2012.09.021.

[312] Zhao, L.; Zhang, Y.; Dai, C.; Guzi, T.; Wiswell, D.; Seghezzi, W.; Parry, D.; Fischmann, T.; Siddiqui, M. A., Design, Synthesis and SAR of Thienopyridines as Potent CHK1 Inhibitors. *Bioorganic Med. Chem. Lett.* **2010**, *20*, 7216–7221. https://doi.org/10.1016/j.bmcl.2010.10.105.

[313] Vorbrüggen, H.; Bennua, B. A., New Simplified Nucleoside Synthesis. *Chem. Ber.* **1981**, *114*, 1279–1286. https://doi.org/10.1002/cber.19811140407.

[314] Antonini, I.; Cristalli, G.; Franchetti, P.; Grifantini, M.; Martelli, S.; Petrelli, F., Deaza Analogues of Adenosine as Inhibitors of Blood Platelet Aggregation. *J. Pharm. Sci.* **1984**, *73* (3), 366–369. https://doi.org/10.1002/jps.2600730319.
[315] Ludwig, J. A., New Route to Nucleoside 5'-Triphosphates. *Acta Biochim. Biophys. Acad. Sci. Hung.* **1981**, *16* (3–4), 131–133. https://doi.org/10.1039/C39910001276.
[316] Niedballa, U.; Vorbrueggen, H. A., General Synthesis of N-Glycosides. 6. On the Mechanism of the Stannic Chloride Catalyzed Silyl Hilbert-Johnson Reaction. *J. Org. Chem.* **1976**, *41* (12), 2084–2086. https://doi.org/10.1021/jo00874a002.
[317] Vorbrüggen, H.; Krolikiewicz, K.; Bennua, B., Nucleoside Synthesis with Trimethylsilyl Triflate and Perchlorate as Catalysts. *Chem. Ber.* **1981**, *114*, 1234–1255. https://doi.org/10.1002/cber.19811140404.
[318] Niedballa, U.; Vorbruggen, H. A., General Synthesis of N-Glycosides. I. Synthesis of Pyrimidine Nucleosides. *J. Org. Chem.* **1974**, *39* (25), 3654–3660. https://doi.org/10.1021/jo00939a008.
[319] Boden, N.; Bushby, R. J.; Cammidge, A. N.; Headdock, G., Iodine Monochloride Can Act as a Chlorinating Agent. *Tetrahedron Lett.* **1995**, *36* (47), 8685–8686. https://doi.org/10.1016/0040-4039(95)01793-H.
[320] Hubig, S. M.; Jung, W.; Kochi, J. K., Cation Radicals as Intermediates in Aromatic Halogenation with Iodine Monochloride: Solvent and Salt Effects on the Competition between Chlorination and Iodination. *J. Org. Chem.* **1994**, *59* (21), 6233–6244. https://doi.org/10.1021/jo00100a025.
[321] Emmanuvel, L.; Shukla, R. K.; Sudalai, A.; Gurunath, S.; Sivaram, S., NaIO4/KI/NaCl: A New Reagent System for Iodination of Activated Aromatics through in Situ Generation of Iodine Monochloride. *Tetrahedron Lett.* **2006**, *47*, 4793–4796. https://doi.org/10.1016/j.tetlet.2006.05.062.
[322] Johnsson, R.; Meijer, A.; Ellervik, U., Mild and Efficient Direct Aromatic Iodination. *Tetrahedron* **2005**, *61*, 11657–11663. https://doi.org/10.1016/j.tet.2005.09.051.
[323] Cava, M. P.; Levinson, M. I., Thionation Reactions of Lawesson's Reagents. *Tetrahedron* **1985**, *41* (22), 5061–5087. https://doi.org/10.1016/S0040-4020(01)96753-5.
[324] Ozturk, T.; Ertas, E.; Mert, O., Use of Lawesson's Reagent in Organic Syntheses. *Chem. Rev.* **2007**, *107* (11), 5210–5278. https://doi.org/10.1021/cr040650b.
[325] Chinchilla, R.; Nájera, C., The Sonogashira Reaction: A Booming Methodology in Synthetic Organic Chemistry. *Chem. Rev.* **2007**, *107*, 874–922. https://doi.org/10.1021/cr050992x.
[326] Hocek, M.; Fojta, M., Cross-Coupling Reactions of Nucleoside Triphosphates Followed by Polymerase Incorporation. Construction and Applications of Base-Functionalized Nucleic Acids. *Org. Biomol. Chem.* **2008**, *6*, 2233–2241. https://doi.org/10.1039/b803664k.
[327] Domnick, C., Expanding the Toolbox of Modified Nucleobases and Click Chemistry for Investigations on Non-Coding RNA, *PhD Thesis,* University of Bonn **2018**.
[328] Hollenstein, M., Nucleoside Triphosphates - Building Blocks for the Modification of Nucleic Acids. *Molecules* **2012**, *17*, 13569–13591. https://doi.org/10.3390/molecules171113569.
[329] Hocek, M., Synthesis of Base-Modified 2'-Deoxyribonucleoside Triphosphates and Their Use in Enzymatic Synthesis of Modified DNA for Applications in Bioanalysis and Chemical Biology. *J. Org. Chem.* **2014**, *79*, 9914–9921. https://doi.org/10.1021/jo5020799.
[330] Yoshikawa, M.; Kato, T.; Takenishi, T., Studies of Phosphorylation. III. Selective Phosphorylation of Unprotected Nucleosides. *Bull. Chem. Soc. Jpn.* **1969**, *42* (12), 3505–3508. https://doi.org/10.1246/bcsj.42.3505.
[331] Someya, T.; Ando, A.; Kimoto, M.; Hirao, I., Site-Specific Labeling of RNA by Combining Genetic Alphabet Expansion Transcription and Copper-Free Click Chemistry. *Nucleic Acids Res.* **2015**, *43* (14), 6665–6676. https://doi.org/10.1093/nar/gkv638.

[332] Ledbetter, M. P.; Malyshev, D. A.; Romesberg, F. E., Site-Specific Labeling of DNA via PCR with an Expanded Genetic Alphabet. In *Practical Research Planning and Design*; 2019; Vol. 1973, pp 193–212. https://doi.org/10.1007/978-1-4939-9216-4_13.
[333] Seo, Y. J.; Malyshev, D. A.; Lavergne, T.; Ordoukhanian, P.; Romesberg, F. E., SI:Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* **2011**, *133* (49), 19878–19888. https://doi.org/10.1021/ja207907d.
[334] Ludwig, J.; Eckstein, F., Rapid and Efficient Synthesis of Nucleoside 5'-O-(1-Thiotriphosphates), 5'-Triphosphates and 2',3'-Cyclophosphorothioates Using 2-Chloro-4H-1,3,2-Benzodioxaphosphorin-4-One. *J. Org. Chem.* **1989**, *54* (3), 631–635. https://doi.org/10.1021/jo00264a024.
[335] Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C., Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates. *Nucleic Acids Res.* **1987**, *15* (21), 8783–8798. https://doi.org/10.1093/nar/15.21.8783.
[336] Domnick, C., Synthese Neuer Tetrazin-Fluorophor Konjugate Für in Vivo Click an Norbornen Modifizierten Oligonukleotiden, *Diploma Thesis,* Univerity of Bonn **2013**.
[337] Domnick, C.; Eggert, F.; Kath-Schorr, S., Site-Specific Enzymatic Introduction of a Norbornene Modified Unnatural Base into RNA and Application in Post-Transcriptional Labeling. *Chem. Commun.* **2015**, *51* (39), 8253–8256. https://doi.org/10.1039/C5CC01765C.
[338] Jia, Y.; Patel, S. S., Kinetic Mechanism of Transcription Initiation by Bacteriophage T7 RNA Polymerase. *Biochemistry* **1997**, *36* (14), 4223–4232. https://doi.org/10.1021/bi9630467.
[339] Uhlenbeck, O. C., Keeping RNA Happy. *RNA* **1995**, *1* (1), 4–6.
[340] Herschlag, D., RNA Chaperones and the RNA Folding Problem. *J. Biol. Chem.* **1995**, *270* (36), 20871–20874. https://doi.org/10.1074/jbc.270.36.20871.
[341] Pan, J.; Thirumalai, D.; Woodson, S. A. ,Folding of RNA Involves Parallel Pathways. *J. Mol. Biol.* **1997**, *273* (1), 7–13. https://doi.org/10.1006/jmbi.1997.1311.
[342] Russel, R., *Biophysics of RNA Folding*; 2013. https://doi.org/10.1007/978-1-4614-8548-3.
[343] Kulikov, K., Charakterisierung Des CPEB3 Ribozyms – Studien Zur Aufklärung Des Mechanismus Eines Humanen Ribozyms, *PhD Thesis,* University of Bonn **2018**.
[344] Eggert, F.; Kulikov, K.; Domnick, C.; Leifels, P.; Kath-Schorr, S., Illuminated by Foreign Letters – Strategies for Site-Specific Cyclopropene Modification of Large Functional RNAs via in Vitro Transcription. *Methods* **2017**, *120*. https://doi.org/10.1016/j.ymeth.2017.04.021.
[345] Pereira, M. J. B.; Harris, D. A.; Rueda, D.; Walter, N. G., Reaction Pathway of the Trans-Acting Hepatitis Delta Virus Ribozyme: A Conformational Change Accompanies Catalysis. *Biochemistry* **2002**, *41*, 730–740. https://doi.org/10.1021/bi011963t.
[346] Chadalavada, D. M.; Gratton, E. A.; Bevilacqua, P. C., The Human HDV-like CPEB3 Ribozyme Is Intrinsically Fast-Reacting. *Biochemistry* **2010**, *49* (25), 5321–5330. https://doi.org/10.1021/bi100434c.
[347] Reiss, C. W.; Xiong, Y.; Strobel, S. A., Structural Basis for Ligand Binding to the Guanidine-I Riboswitch. *Structure* **2017**, *25*, 195–202. https://doi.org/10.1016/j.str.2016.11.020.
[348] Shevchuk, N. A.; Bryksin, A. V.; Nusinovich, Y. A.; Cabello, F. C.; Sutherland, M.; Ladisch, S., Construction of Long DNA Molecules Using Long PCR-Based Fusion of Several Fragments Simultaneously. *Nucleic Acids Res.* **2004**, *32* (2), e19. https://doi.org/10.1093/nar/gnh014.
[349] Potier, N.; Van Dorsselaer, A.; Cordier, Y.; Roch, O.; Bischoff, R., Negative Electrospray Ionization Mass Spectrometry of Synthetic and Chemically Modified Oligonucleotides. *Nucleic Acids Res.* **1994**, *22* (19), 3895–3903. https://doi.org/10.1093/nar/22.19.3895.
[350] Basiri, B.; Bartlett, M. G., LC–MS of Oligonucleotides: Applications in Biomedical Research. *Bioanalysis* **2014**, *6* (11), 1525–1542. https://doi.org/10.4155/bio.14.94.

[351] Hail, M. E.; Elliot, B.; Anderson, K., High-Throughput Analysis of Oligonucleotides Using Automated Electrospray Ionization Mass Spectrometry. *Am. Biotechnol. Lab.* **2004**, *22*, 12–14.
[352] Driegen, S.; Ferreira, R.; Van Zon, A.; Strouboulis, J.; Jaegle, M.; Grosveld, F.; Philipsen, S.; Meijer, D. A., Generic Tool for Biotinylation of Tagged Proteins in Transgenic Mice. *Transgenic Res.* **2005**, *14* (4), 477–482. https://doi.org/10.1007/s11248-005-7220-2.
[353] Mercer, T. R.; Gerhardt, D. J.; Dinger, M. E.; Crawford, J.; Trapnell, C.; Jeddeloh, J. A.; Mattick, J. S.; Rinn, J. L., Targeted RNA Sequencing Reveals the Deep Complexity of the Human Transcriptome. *Nat. Biotechnol.* **2012**, *30* (1), 99–104. https://doi.org/10.1038/nbt.2024.
[354] Mercer, T. R.; Dinger, M. E.; Mattick, J. S., Long Non-Coding RNAs: Insights into Functions. *Nat. Rev. Genet.* **2009**, *10* (3), 155–159. https://doi.org/10.1038/nrg2521.
[355] Stellwagen, N. C. Electrophoresis of DNA in Agarose Gels, Polyacrylamide Gels and in Free Solution. *Electrophoresis* **2009**, *30* (S1), S188–S195. https://doi.org/10.1002/elps.200900052.
[356] Pereira, M. J. B.; Behera, V.; Walter, N. G., Nondenaturing Purification of Co-Transcriptionally Folded RNA Avoids Common Folding Heterogeneity. *PLoS One* **2010**, *5* (9), e12953. https://doi.org/10.1371/journal.pone.0012953.
[357] Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A., Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjug. Chem.* **2008**, *19*, 2297–2299. https://doi.org/10.1021/bc8004446.
[358] Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y., Diels–Alder Cycloaddition for Fluorophore Targeting to Specific Proteins inside Living Cells. *J. Am. Chem. Soc.* **2012**, *134* (2), 792–795. https://doi.org/10.1021/ja209325n.
[359] Šečkute, J.; Devaraj, N. K., Expanding Room for Tetrazine Ligations in the in Vivo Chemistry Toolbox. *Curr. Opin. Chem. Biol.* **2013**, *17*, 761–767. https://doi.org/10.1016/j.cbpa.2013.08.004.
[360] Prentki, P.; Krisch, H. M., In Vitro Insertional Mutagenesis with a Selectable DNA Fragment. *Gene* **1984**, *29* (3), 303–313. https://doi.org/10.1016/0378-1119(84)90059-3.
[361] Muyrers, J. P. .; Zhang, Y.; Stewart, A. F., Techniques: Recombinogenic Engineering–New Options for Cloning and Manipulating DNA. *Trends Biochem. Sci.* **2001**, *26* (5), 325–331. https://doi.org/10.1016/S0968-0004(00)01757-6.
[362] Baumann, P., *G-Quadruplex DNA Methods and Protocols*, 1st ed.; Walker, J. M., Ed.; Heidelberg, 2010.
[363] Feldman, A. W.; Romesberg, F. E., Expansion of the Genetic Alphabet: A Chemist's Approach to Synthetic Biology. *Acc. Chem. Res.* **2018**, *51*, 394–403. https://doi.org/10.1021/acs.accounts.7b00403.
[364] Cheetham, G. M. T.; Steitz, T. A., Structure of a Transcribing T7 RNA Polymerase Initiation Complex. *Science* **1999**, *286* (5448), 2305–2309. https://doi.org/10.1126/science.286.5448.2305.
[365] Yin, Y. W.; Steitz, T. A., Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase. *Science* **2002**, *298*, 1387–1395. https://doi.org/10.1126/science.1077464.
[366] Pleiss, J. A.; Derrick, M. L.; Uhlenbeck, O. C., T7 RNA Polymerase Produces 5′ End Heterogeneity during in Vitro Transcription from Certain Templates. *RNA* **1998**, *4*, 1313–1317. https://doi.org/10.1017/S135583829800106X.
[367] Doudna, J. A.; Cech, T. R., The Chemical Repertoire of Natural Ribozymes. *Nature* **2002**, *418*, 222–228. https://doi.org/10.1038/418222a.
[368] Kao, C.; Zheng, M.; Rüdisser, S. A., Simple and Efficient Method to Reduce Nontemplated Nucelotide Addition at the 3 Term of RNAs Transcribed by T7 RNA Polymerase. *RNA*. Cambridge University Press 1999, pp 1268–1272.
[369] Kao, C. C., Reducing Nontemplated 3' Nucleotide Addition To Polynucleotide Transcripts. US6090590A, 2000.

[370] Eggert, F.; Kurscheidt, K.; Hoffmann, E.; Kath-Schorr, S., Towards Reverse Transcription with an Expanded Genetic Alphabet. *ChemBioChem* **2019**, 1642–1645. https://doi.org/10.1002/cbic.201800808.
[371] Hoffmann, E., Transcription and Reverse Transcription of an Expanded Genetic Alphabet, University of Bonn **2019**.
[372] Kurscheidt, K. Reverse Transcription of Unnatural Base Pairs, *Bachelor Thesis,* University of Bonn **2018**.
[373] Mahara, A.; Iwase, R.; Sakamoto, T.; Yamana, K.; Yamaoka, T.; Murakami, A., Bispyrene-Conjugated 2′-O-Methyloligonucleotide as a Highly Specific RNA-Recognition Probe. *Angew. Chemie Int. Ed.* **2002**, *41* (19), 3648–3650. https://doi.org/10.1002/1521-3773(20021004)41:19<3648::AID-ANIE3648>3.0.CO;2-Y.
[374] Yin, D. M.; Hannam, J. S.; Schmitz, A.; Schiemann, O.; Hagelueken, G.; Famulok, M., Studying the Conformation of a Receptor Tyrosine Kinase in Solution by Inhibitor-Based Spin Labeling. *Angew. Chemie - Int. Ed.* **2017**, *129*, 8537–8541. https://doi.org/10.1002/anie.201703154.
[375] Domnick, C.; Hagelueken, G.; Eggert, F.; Schiemann, O.; Kath-Schorr, S., Posttranscriptional Spin Labeling of RNA by Tetrazine-Based Cycloaddition. *Org. Biomol. Chem.* **2019**, *17* (7), 1805–1808. https://doi.org/10.1039/C8OB02597E.
[376] Eggert, F.; Kath-Schorr, S., A Cyclopropene-Modified Nucleotide for Site-Specific RNA Labeling Using Genetic Alphabet Expansion Transcription. *Chem. Commun.* **2016**, *52*, 7284–7287. https://doi.org/10.1039/C6CC02321E.
[377] Marky, L. A.; Blumenfeld, K. S.; Kozlowski, S.; Breslauer, K. J., Salt-Dependent Conformational Transitions in the Self-Complementary Deoxydodecanucleotides d(CGCGCAATTCGCG): Evidence for Hairpin Formation. *Biopolymers* **1983**, *22*, 1247–1257.
[378] Kirchner, R.; Vogtherr, M.; Limmer, S.; Sprinzl, M., Secondary Structure Dimorphism and Interconversion Between Hairpin and Duplex Form of Oligoribonucleotides. *Antisense Nucleic Acid Drug Dev.* **1998**, *8*, 507–516. https://doi.org/10.1089/oli.1.1998.8.507.
[379] Nakano, S.; Kirihata, T.; Fujii, S.; Sakai, H.; Kuwahara, M.; Sawai, H.; Sugimoto, N., Influence of Cationic Molecules on the Hairpin to Duplex Equilibria of Self-Complementary DNA and RNA Oligonucleotides. *Nucleic Acids Res.* **2007**, *35* (2), 486–494. https://doi.org/10.1093/nar/gkl1073.
[380] Hagelueken, G.; Ward, R.; Naismith, J. H.; Schiemann, O., MtsslWizard: In Silico Spin-Labeling and Generation of Distance Distributions in PyMOL. *Appl. Magn. Reson.* **2012**, *42*, 377–391. https://doi.org/10.1007/s00723-012-0314-0.
[381] Vorlíčková, M.; Kejnovská, I.; Bednářová, K.; Renčiuk, D.; Kypr, J., Circular Dichroism Spectroscopy of DNA: From Duplexes to Quadruplexes. *Chirality* **2012**, *24* (9), 691–698. https://doi.org/10.1002/chir.22064.
[382] Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M., Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725. https://doi.org/10.1093/nar/gkp026.
[383] Sicoli, G.; Wachowius, F.; Bennati, M.; Höbartner, C., Probing Secondary Structures of Spin-Labeled RNA by Pulsed EPR Spectroscopy. *Angew. Chemie Int. Ed.* **2010**, *49* (36), 6443–6447. https://doi.org/10.1002/anie.201000713.
[384] Http://Biotools.Nubic.Northwestern.Edu/OligoCalc.Html.
[385] Zhang, W.; Chen, S.-J., Exploring the Complex Folding Kinetics of RNA Hairpins: II. Effect of Sequence, Length, and Misfolded States. *Biophys. J.* **2006**, *90* (3), 778–787. https://doi.org/10.1529/biophysj.105.062950.
[386] Raman, E. P.; Lemkul, J. A.; Best, R.; Mackerell, A. D. E., Prabhu Raman, Justin A. Lemkul, Robert Best, Alexander D. MacKerell, Jr., CHARMM36 Port: Http://Mackerell.Umaryland.Edu/Charmm_ff.Shtml. 36.
[387] Pley, H. W.; Flaherty, K. M.; McKay, D. B., Three-Dimensional Structure of a Hammerhead Ribozyme. *Nature* **1994**, *372* (6501), 68–74. https://doi.org/10.1038/372068a0.

[388] Birikh, K. R.; Heaton, P. A.; Eckstein, F., The Structure, Function and Application of the Hammerhead Ribozyme. *Eur. J. Biochem.* **1997**, *245* (1), 1–16. https://doi.org/10.1111/j.1432-1033.1997.t01-3-00001.x.
[389] Vaish, N. K.; Kore, A. R.; Eckstein, F., Recent Developments in the Hammerhead Ribozyme Field. *Nucleic Acids Res.* **1998**, *26* (23), 5237–5242. https://doi.org/10.1093/nar/26.23.5237.
[390] Rydzanicz, R.; Zhao, X. S.; Johnson, P. E., Assembly PCR Oligo Maker: A Tool for Designing Oligodeoxynucleotides for Constructing Long DNA Molecules for RNA Production. *Nucleic Acids Res.* **2005**, *33* (Web Server), W521–W525. https://doi.org/10.1093/nar/gki380.
[391] Krupp, G., RNA Synthesis: Strategies for the Use of Bacteriophage RNA Polymerases. *Gene* **1988**, *72* (1–2), 75–89. https://doi.org/10.1016/0378-1119(88)90129-1.
[392] Ashwood, B.; Jockusch, S.; Crespo-Hernández, C. E., Photochemical Reactivity of DTPT3: A Crucial Nucleobase Derivative in the Development of Semisynthetic Organisms. *J. Phys. Chem. Lett.* **2017**, *8* (11), 2387–2392. https://doi.org/10.1021/acs.jpclett.7b00926.
[393] Pollum, M.; Ashwood, B.; Jockusch, S.; Lam, M.; Crespo-Hernández, C. E., Unintended Consequences of Expanding the Genetic Alphabet. *J. Am. Chem. Soc.* **2016**, *138* (36), 11457–11460. https://doi.org/10.1021/jacs.6b06822.
[394] Ashwood, B.; Pollum, M.; Crespo-Hernández, C. E., Photochemical and Photodynamical Properties of Sulfur-Substituted Nucleic Acid Bases,. *Photochem. Photobiol.* **2019**, *95* (1), 33–58. https://doi.org/10.1111/php.12975.
[395] Wang, Q.; Xie, X.-Y.; Han, J.; Cui, G., QM and QM/MM Studies on Excited-State Relaxation Mechanisms of Unnatural Bases in Vacuo and Base Pairs in DNA. *J. Phys. Chem. B* **2017**, *121* (46), 10467–10478. https://doi.org/10.1021/acs.jpcb.7b09046.
[396] Chen, Z.; Zhang, Y., Dimethyl Sulfoxide Targets Phage RNA Polymerases to Promote Transcription. *Biochem. Biophys. Res. Commun.* **2005**, *333*, 664–670. https://doi.org/10.1016/j.bbrc.2005.05.166.
[397] Comins, D. L.; Killpack, M. O., Lithiation of Heterocycles Directed by A-Amino Alkoxides. *J. Org. Chem.* **1987**, *52*, 104–109. https://doi.org/10.1021/jo00377a019.
[398] Hornberger, K. R.; Berger, D. M.; Chen, X.; Crew, A. P.; Dong, H.; Kleinberg, A.; Li, A.-H.; Ma, L.; Mulcihill, M. J.; Panicker, B.; et al., 7-Aminofuropyridine Derivatives, 2013. https://doi.org/10.1016/j.(73).
[399] Heck, R. F., Palladium-Catalyzed Reactions of Organic Halides with Olefins. *Acc. Chem. Res.* **1979**, *12* (4), 146–151. https://doi.org/10.1021/ar50136a006.
[400] Beletskaya, I. P.; Cheprakov, A. V., The Heck Reaction as a Sharpening Stone of Palladium Catalysis. *Chem. Rev.* **2000**, *100* (8), 3009–3066. https://doi.org/10.1021/cr9903048.
[401] Hartung, C. G.; Köhler, K.; Beller, M., Highly Selective Palladium-Catalyzed Heck Reactions of Aryl Bromides with Cycloalkenes. *Org. Lett.* **1999**, *1* (5), 709–711. https://doi.org/10.1021/ol9901063.
[402] Wu, X.; Zhou, J., Selective Arylation at the Vinylic Site of Cyclic Olefins. *Chem. Commun.* **2013**, *49*, 4794–4796. https://doi.org/10.1039/c3cc41722k.
[403] Willemse, T.; Schepens, W.; Vlijmen, H.; Maes, B.; Ballet, S., The Suzuki–Miyaura Cross-Coupling as a Versatile Tool for Peptide Diversification and Cyclization. *Catalysts* **2017**, *7* (12), 74. https://doi.org/10.3390/catal7030074.
[404] Buchspies, J.; Szostak, M., Recent Advances in Acyl Suzuki Cross-Coupling. *Catalysts* **2019**, *9* (1), 53. https://doi.org/10.3390/catal9010053.
[405] Pijper, T. C.; Kudernac, T.; Browne, W. R.; Feringa, B. L., Effect of Immobilization on Gold on the Temperature Dependence of Photochromic Switching of Dithienylethenes. *J. Phys. Chem. C* **2013**, *117* (34), 17623–17632. https://doi.org/10.1021/jp404925m.
[406] Miyaura, N.; Suzuki, A., Palladium-Catalyzed Cross-Coupling Reactions of Organoboron Compounds. *Chem. Rev.* **1995**, *95* (7), 2457–2483. https://doi.org/10.1021/cr00039a007.

[407] Franzén, R.; Xu, Y., Review on Green Chemistry - Suzuki Cross Coupling in Aqueous Media. *Can. J. Chem.* **2005**, *83* (3), 266–272. https://doi.org/10.1139/v05-048.
[408] Lucas, L. N.; de Jong, J. J. D.; van Esch, J. H.; Kellogg, R. M.; Feringa, B. L., Syntheses of Dithienylcyclopentene Optical Molecular Switches. *Eur. J. Org. Chem.* **2003**, 155–166. https://doi.org/10.1002/1099-0690.
[409] Jong, J. J. D. de; Lucas, L. N.; Hania, R.; Pugzlys, A.; Kellogg, R. M.; Feringa, B. L.; Duppen, K.; Esch, J. H. van, Photochromic Properties of Perhydro- and Perfluorodithienylcyclopentene Molecular Switches. *European J. Org. Chem.* **2003**, *2003* (10), 1887–1893. https://doi.org/10.1002/ejoc.200200719.
[410] Li, J. J.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Collins, J. T.; Garland, D. J.; Gregory, S. A.; Huang, H.-C.; Isakson, P. C., 1,2-Diarylcyclopentenes as Selective Cyclooxygenase-2 Inhibitors and Orally Active Anti-Inflammatory Agents. *J. Med. Chem.* **1995**, *38* (22), 4570–4578. https://doi.org/10.1021/jm00022a023.
[411] Rolland, V.; Kotera, M.; Lhomme, J., Convenient Preparation of 2-Deoxy-3,5-Di-O- p -Toluoyl-α-D- Erythro -Pentofuranosyl Chloride. *Synth. Commun.* **1997**, *27* (20), 3505–3511. https://doi.org/10.1080/00397919708007071.
[412] Gillerman, I.; Fischer, B., An Improved One-Pot Synthesis of Nucleoside 5′-Triphosphate Analogues. *Nucleosides, Nucleotides and Nucleic Acids* **2010**, *29* (3), 245–256. https://doi.org/10.1080/15257771003709569.
[413] Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G., DNA Modification under Mild Conditions by Suzuki-Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew. Chemie Int. Ed.* **2013**, *52* (40), 10553–10558. https://doi.org/10.1002/anie.201304038.
[414] Tolle, F.; Rosenthal, M.; Pfeiffer, F.; Mayer, G., Click Reaction on Solid Phase Enables High Fidelity Synthesis of Nucleobase-Modified DNA. *Bioconjug. Chem.* **2016**, *27*, 500–503. https://doi.org/10.1021/acs.bioconjchem.5b00668.
[415] Zhang, L., *Chapter 3: Unnatural Nucleic Acids for Aptamer Selection*; 2015. https://doi.org/10.1007/978-3-662-46226-3.
[416] Pfeiffer, F.; Rosenthal, M.; Siegl, J.; Ewers, J.; Mayer, G., Customised Nucleic Acid Libraries for Enhanced Aptamer Selection and Performance. *Curr. Opin. Biotechnol.* **2017**, *48*, 111–118. https://doi.org/10.1016/j.copbio.2017.03.026.
[417] Svarovskaia, S. E.; Cheslock, S. R.; Zhang, W.-H.; Hu, W.-S.; Pathak, V. K., Retroviral Mutation Rates And Reverse Transcriptase Fidelity. *Front Biosci* **2003**, *1*, 117–134.
[418] Kranaster, R.; Zeller, J.; Köhn, B.; Marx, A., Neues Enzym Mit Reverse Transkriptase und DNA-Polymerase-Funktion. *BIOspektrum* **2016**, *22* (2), 164–165. https://doi.org/10.1007/s12268-016-0671-6.
[419] Tidwell, T. T., Wilhelm Schlenk: The Man Behind the Flask. *ChemInform* **2001**, *32* (15), no-no. https://doi.org/10.1002/chin.200115263.
[420] Sigma-Aldrich, Technical Bulletin AL-134: Handling Air-Sensitive Reagents. *Tech. Bull.* **2012**, *2016* (800).
[421] Fulmer, G. R.; Miller, A. J. M.; Sherden, N. H.; Gottlieb, H. E.; Nudelman, A.; Stoltz, B. M.; Bercaw, J. E.; Goldberg, K. I., NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **2010**, *29* (9), 2176–2179. https://doi.org/10.1021/om100106e.
[422] Moffatt, J. G., A General Synthesis of Nucleoside-5'-Triphosphates. *Can. J. Chem.* **1964**, *42* (3), 599–604. https://doi.org/10.1139/v64-087.
[423] Hu, Q.-J.; Lu, Y.-C.; Yang, C.-X.; Yan, X.-P., Synthesis of Covalently Bonded Boron-Dipyrromethene–Diarylethene for Building a Stable Photosensitizer with Photo-Controlled Reversibility. *Chem. Commun.* **2016**, *52*, 5470–5473. https://doi.org/10.1039/C6CC01864E.
[424] Irie, M.; Sakemura, K.; Okinaka, M.; Uchida, K., Photochromism of Dithienylethenes with Electron-Donating Substituents. *J. Org. Chem.* **1995**, *60*, 8305–8309. https://doi.org/10.1021/jo00130a035.

[425] Rossi, S.; Benaglia, M.; Cirilli, R.; Benincori, T., Synthesis of Novel Chiral Bithiophenebased Phosphine Oxides as Lewis Bases in Organocatalytic Stereoselective Reactions. *Asymmetric Catal.* **2015**, *2* (1), 17–25. https://doi.org/10.1515/asorg-2015-0002.
[426] Hervé, G.; Len, C. ,Heck and Sonogashira Couplings in Aqueous Media – Application to Unprotected Nucleosides and Nucleotides. *Sustain. Chem. Process.* **2015**, *3* (3), 1–14. https://doi.org/10.1186/s40508-015-0029-2.
[427] Nielsen, H., Working with RNA. In *RNA: Methods and Protocols*; Nielsen, H., Ed.; Humana Press: Totowa, NJ, 2011; pp 15–28. https://doi.org/10.1007/978-1-59745-248-9_2.
[428] Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostel, J.; Wheeler, D. L., GenBank. *Nucleic Acids Res.* **2004**, *33* (Database issue), D34–D38. https://doi.org/10.1093/nar/gki063.
[429] Mulhardt, C., Molecular Biology and Genomics **2007**, Academic Press, 52.

7 Appendix

7.1 NMR spectra

The following section contains all NMR spectra necessary to specify the known and novel compounds described in section 5.2.7. Solvent residual signals (sol.) and assignable impurities are marked with asterisks.

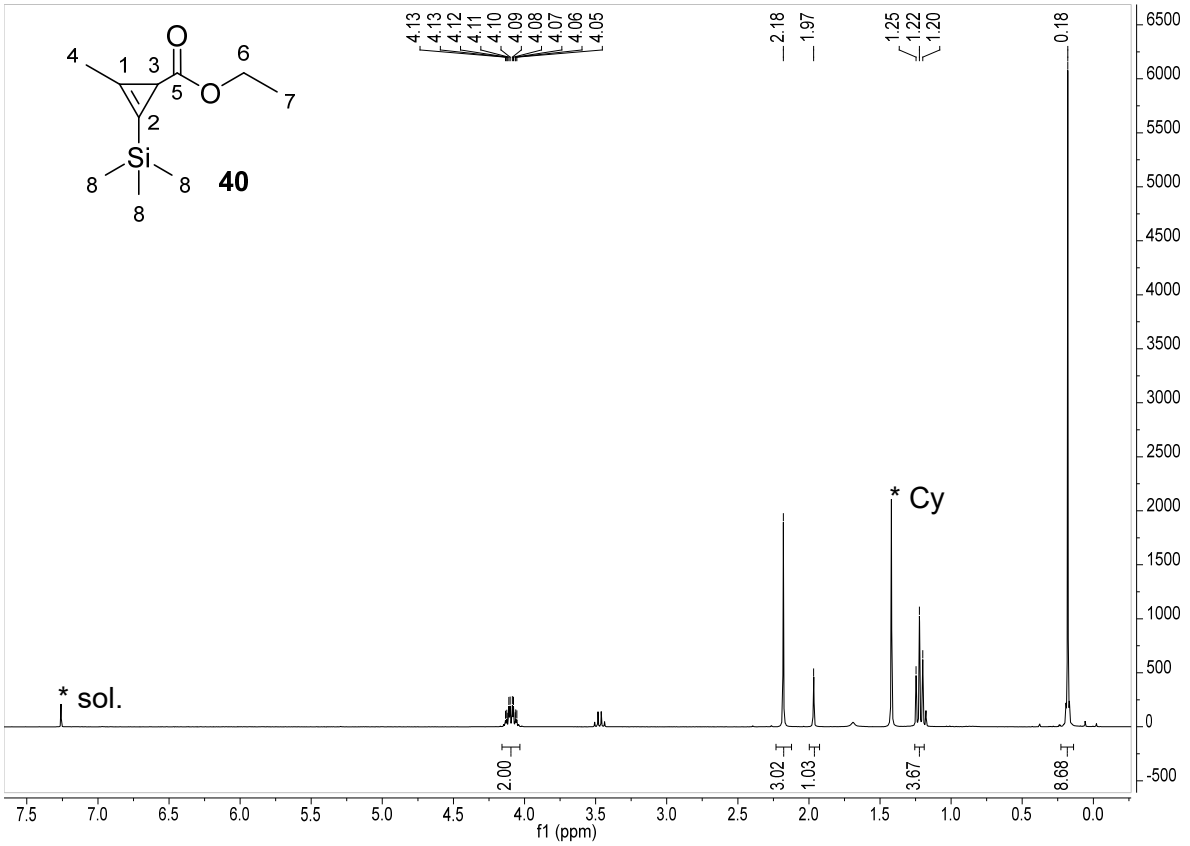


Figure 156. ^{1}H-NMR spectrum ($CDCl_3$, 300 MHz) of cyclopropene ethyl ester **40**.

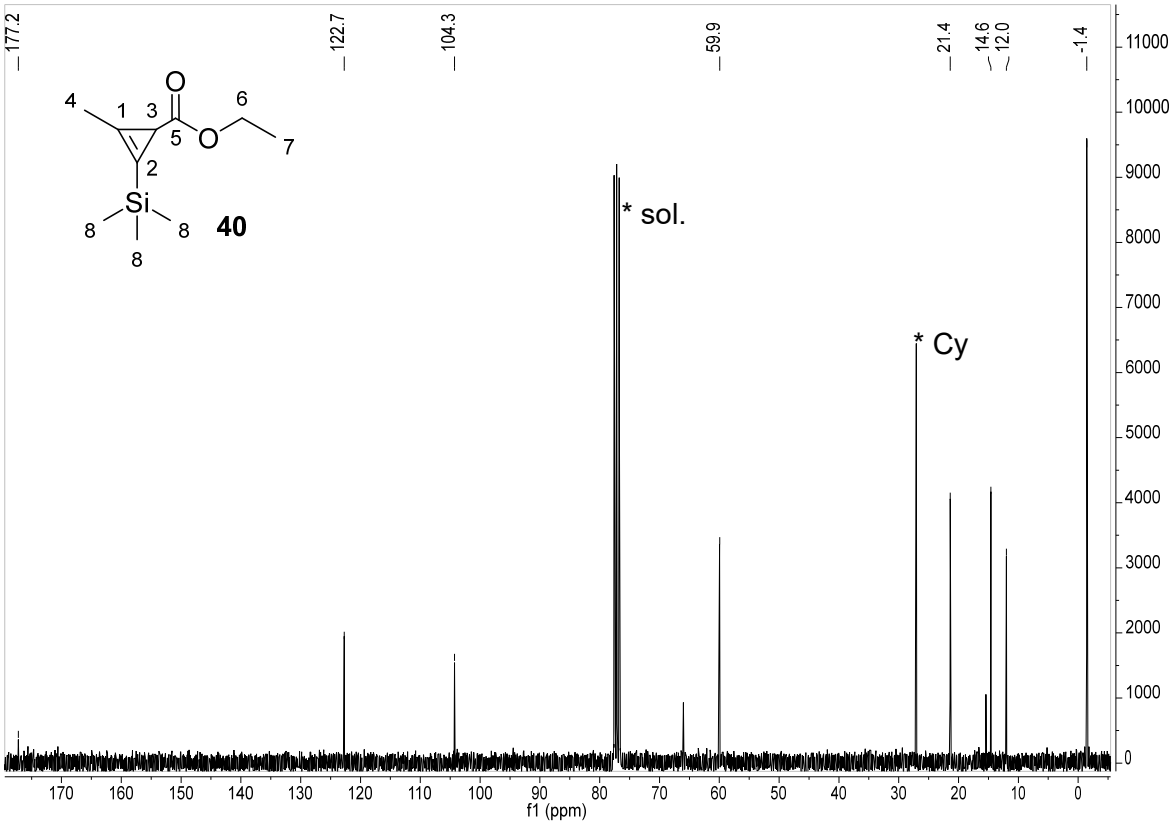


Figure 157. ^{13}C-NMR spectrum ($CDCl_3$, 75 MHz) of cyclopropene ethyl ester **40**.

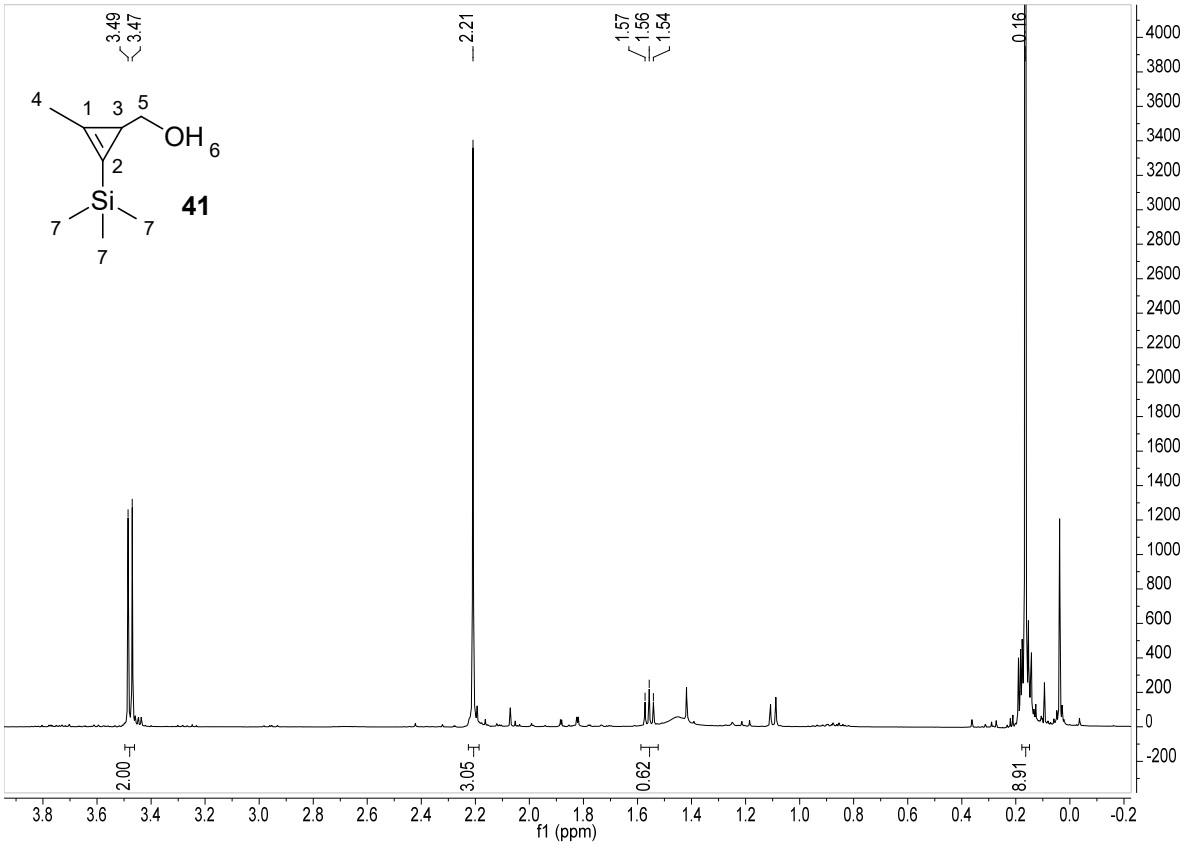


Figure 158. ^{1}H-NMR spectrum ($CDCl_3$, 300 MHz) of cyclopropene alcohol **41**.

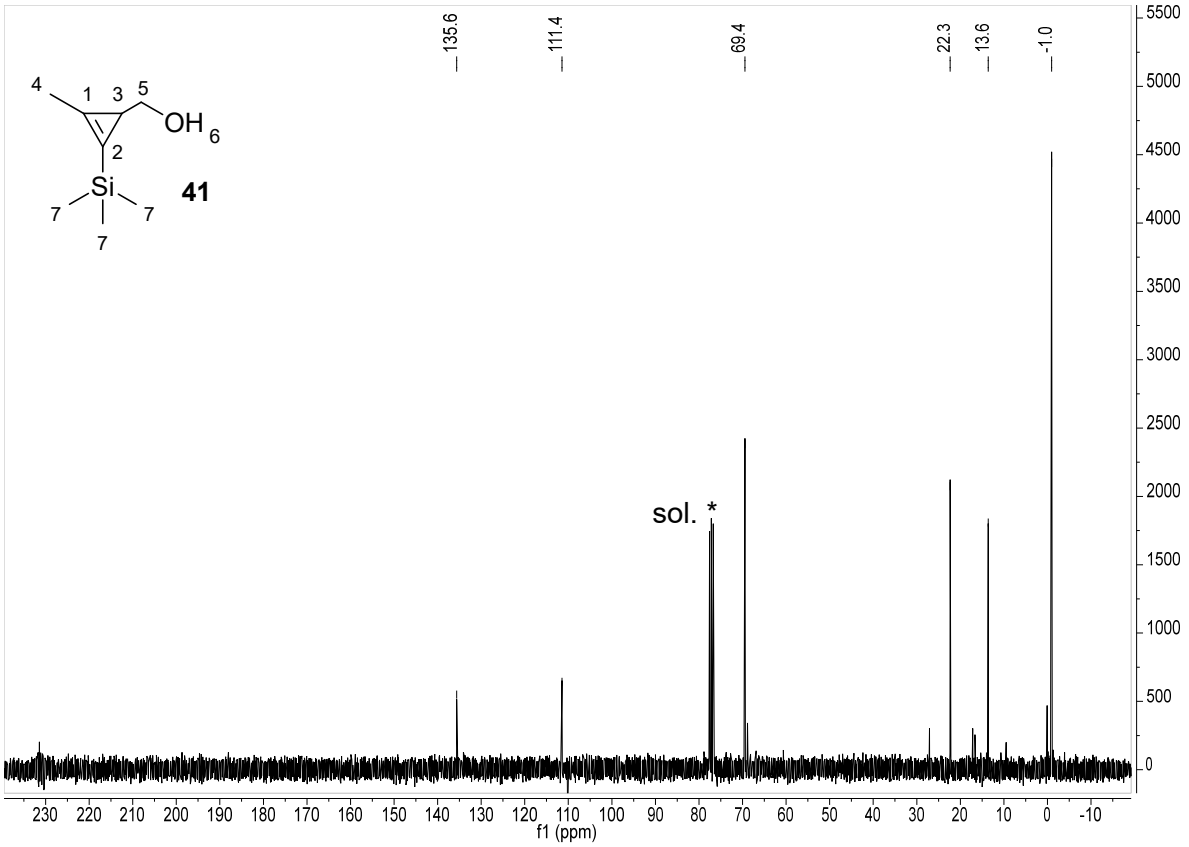


Figure 159. ^{13}C-NMR spectrum ($CDCl_3$, 75 MHz) of cyclopropene alcohol **41**.

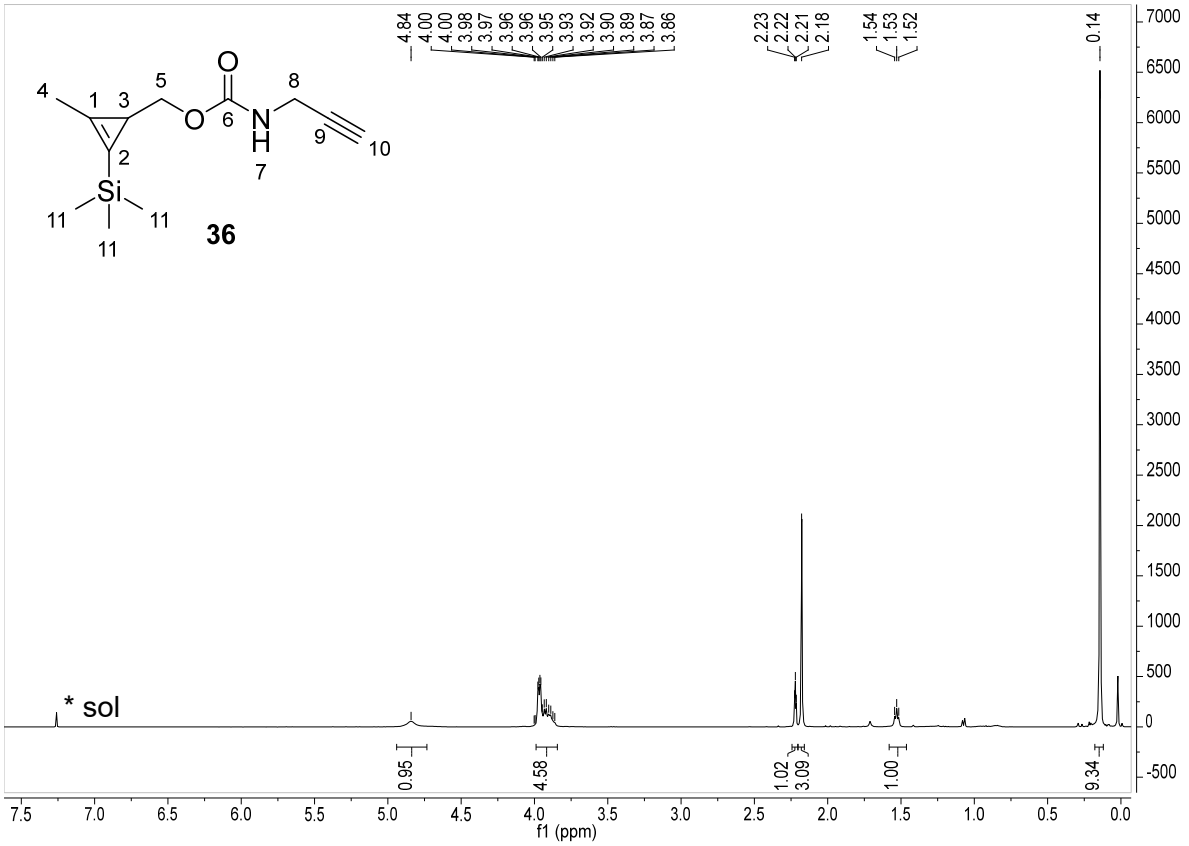


Figure 160. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of carbamate **36**.

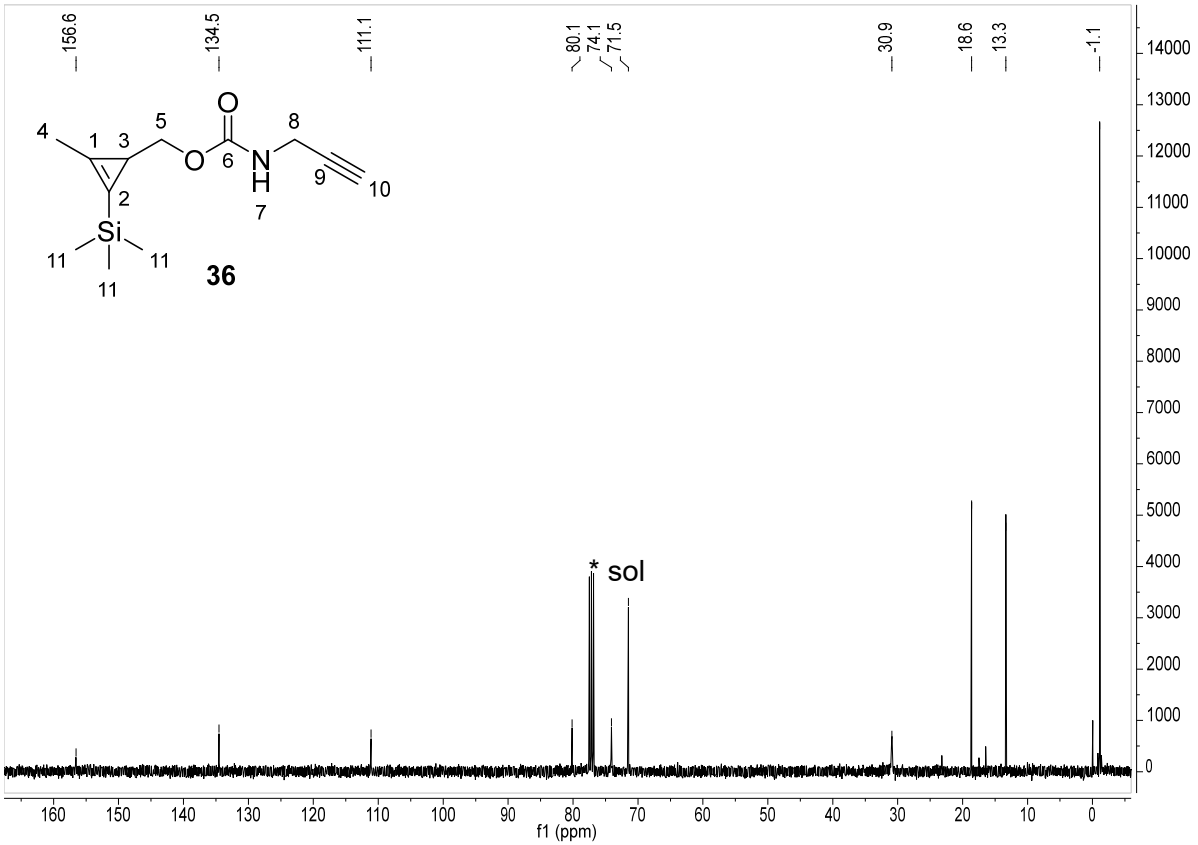


Figure 161. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of carbamate **36**.

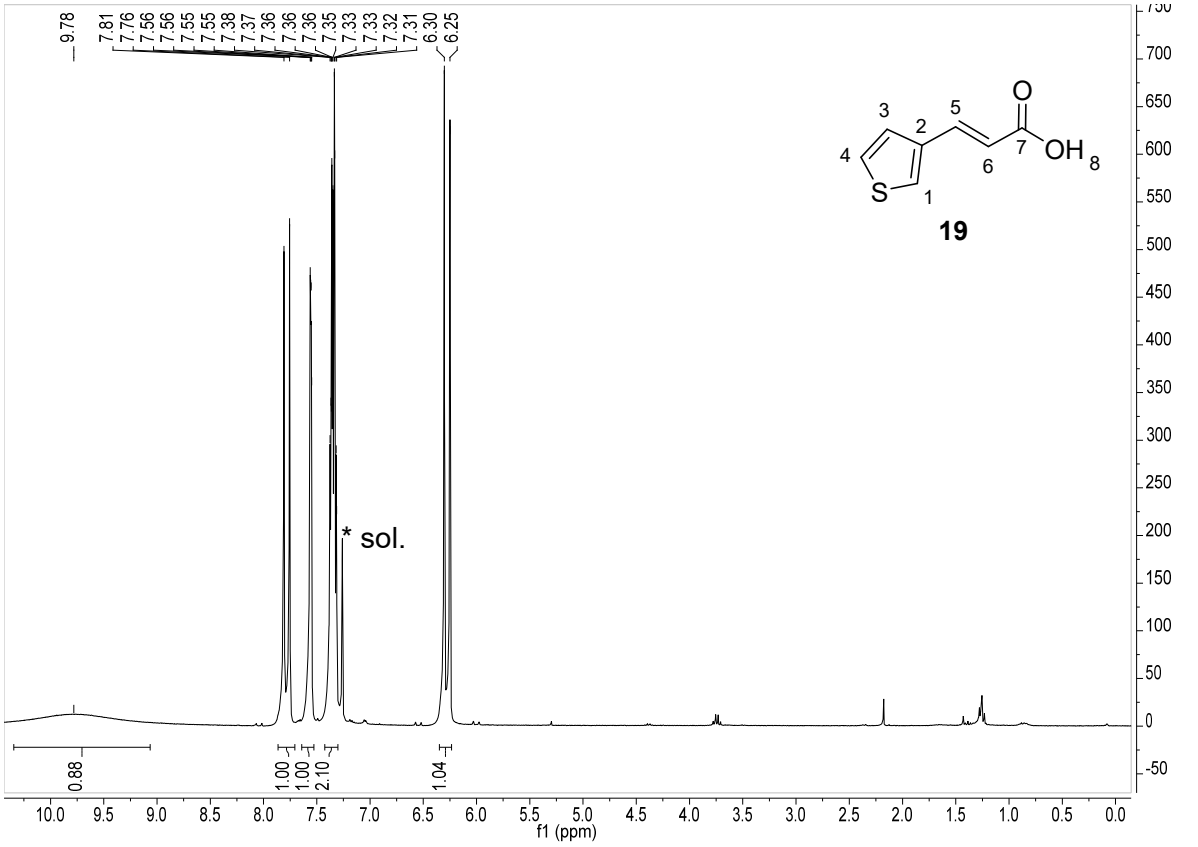


Figure 162. ^{1}H-NMR spectrum ($CDCl_3$, 300 MHz) of acrylic acid **19**.

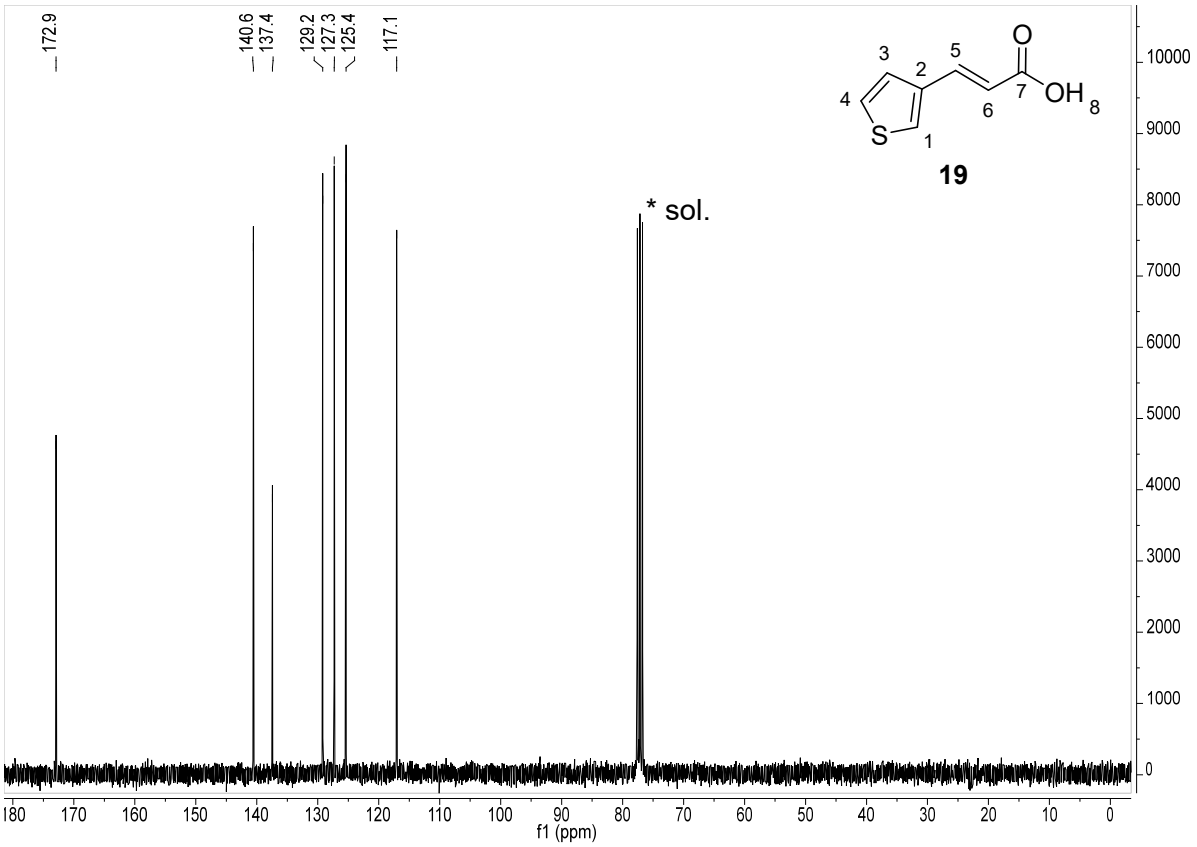


Figure 163. ^{13}C-NMR spectrum ($CDCl_3$, 75 MHz) of acrylic acid **19**.

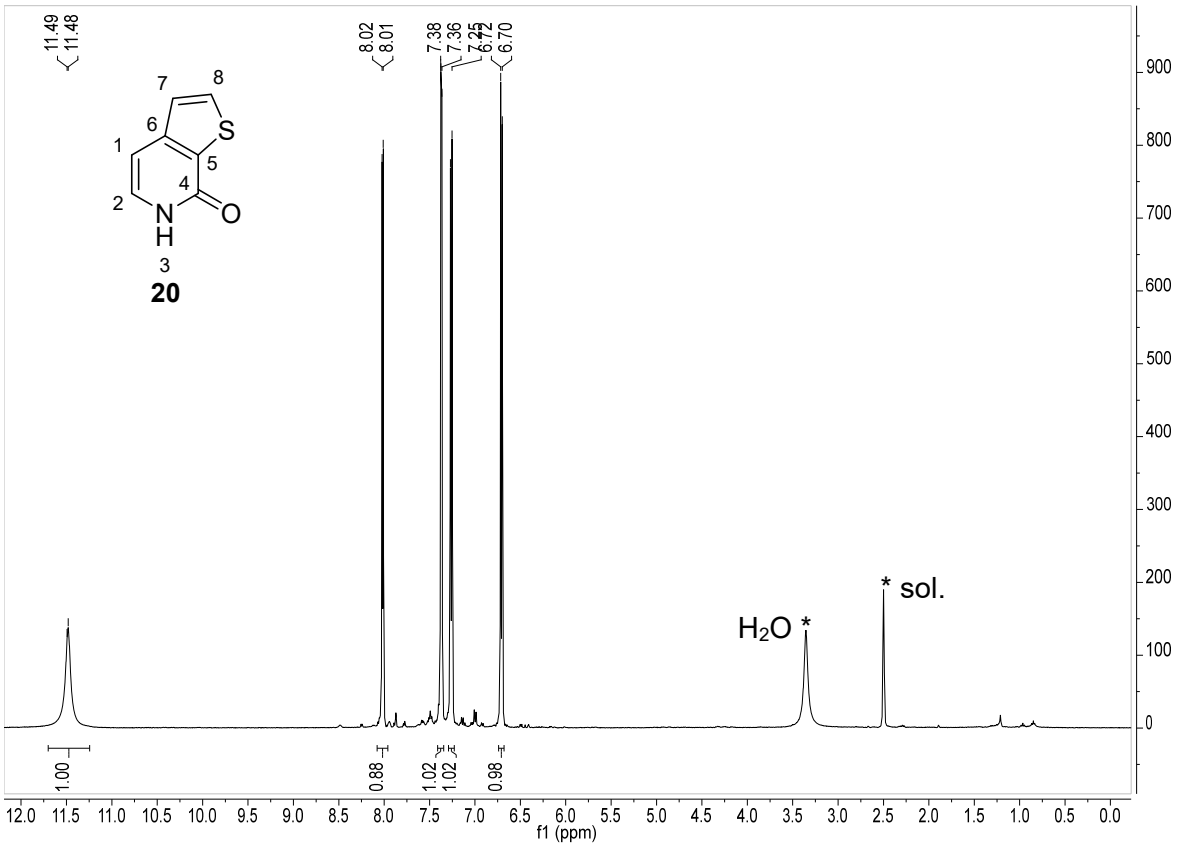


Figure 164. ^{1}H-NMR spectrum (DMSO-d_6, 400 MHz) of pyridone derivative **20**.

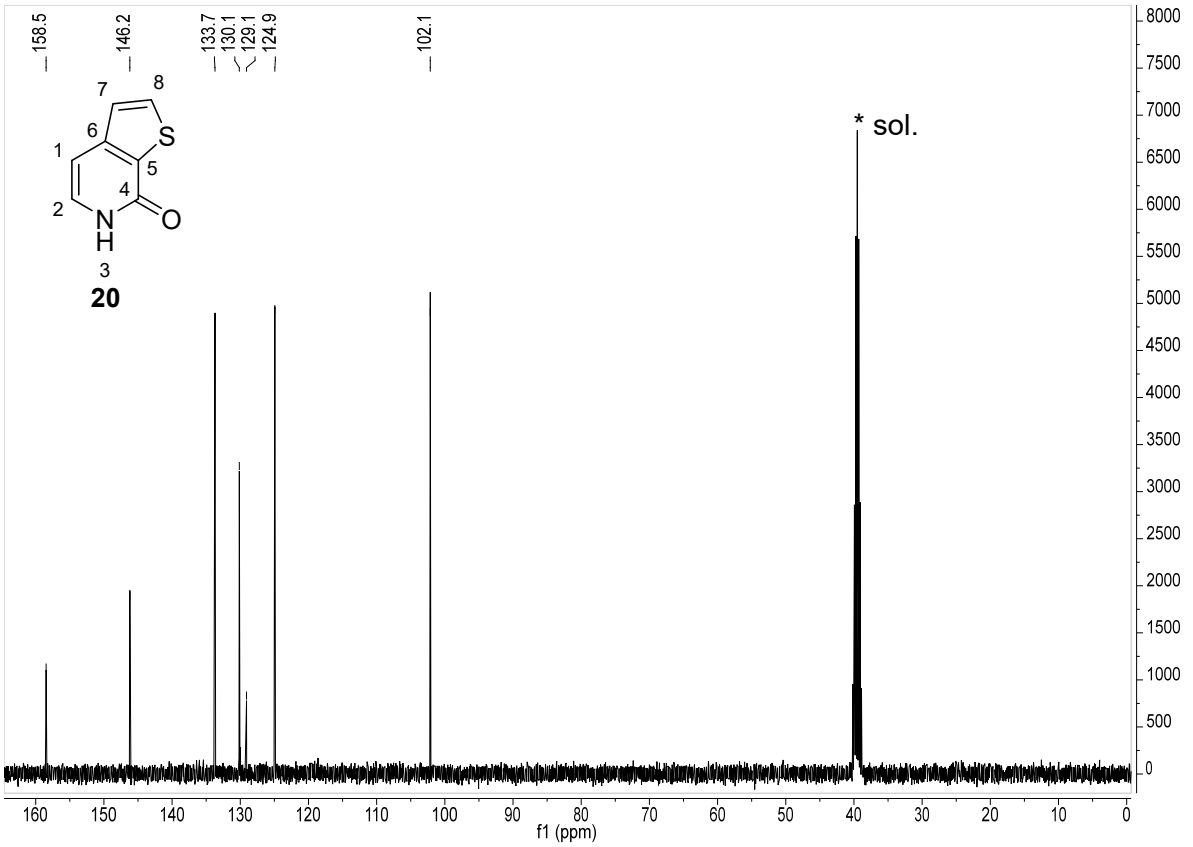


Figure 165. ^{13}C-NMR spectrum (DMSO-d_6, 101 MHz) of pyridone derivative **20**.

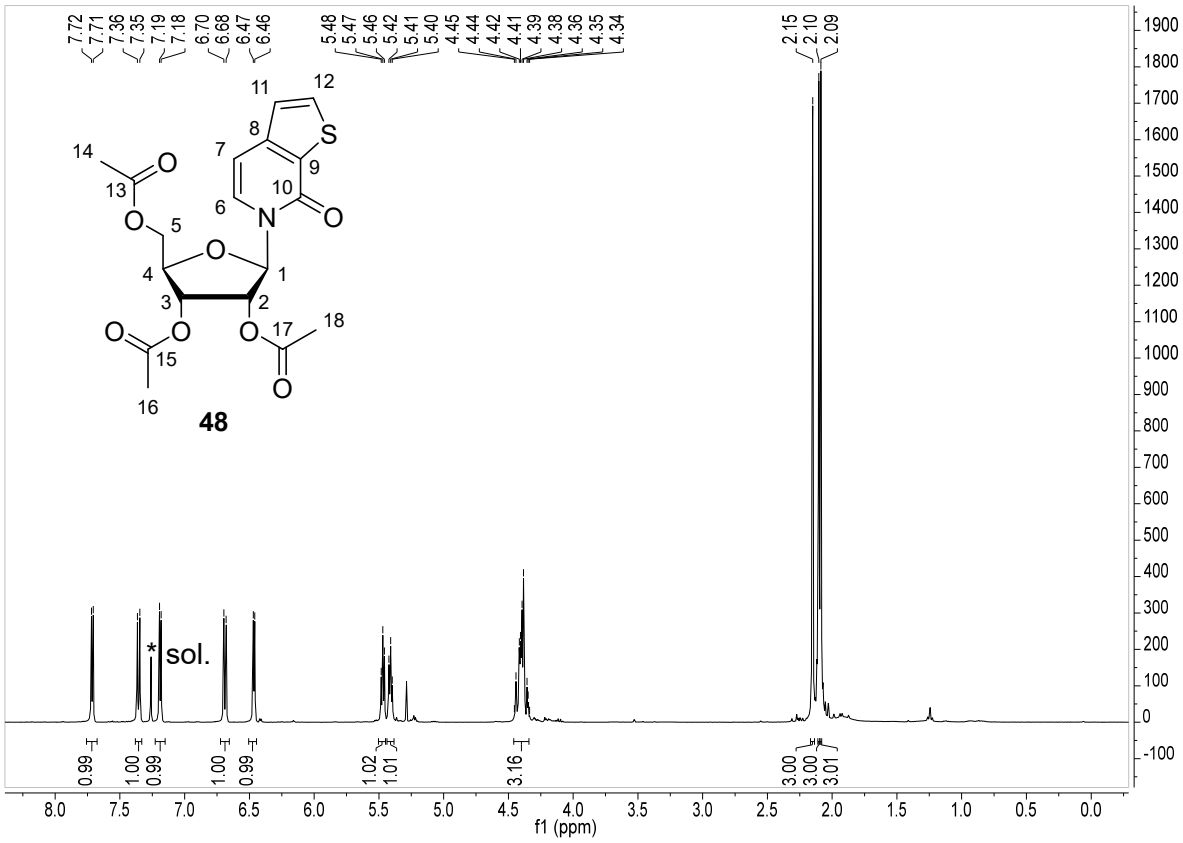


Figure 166. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **48**.

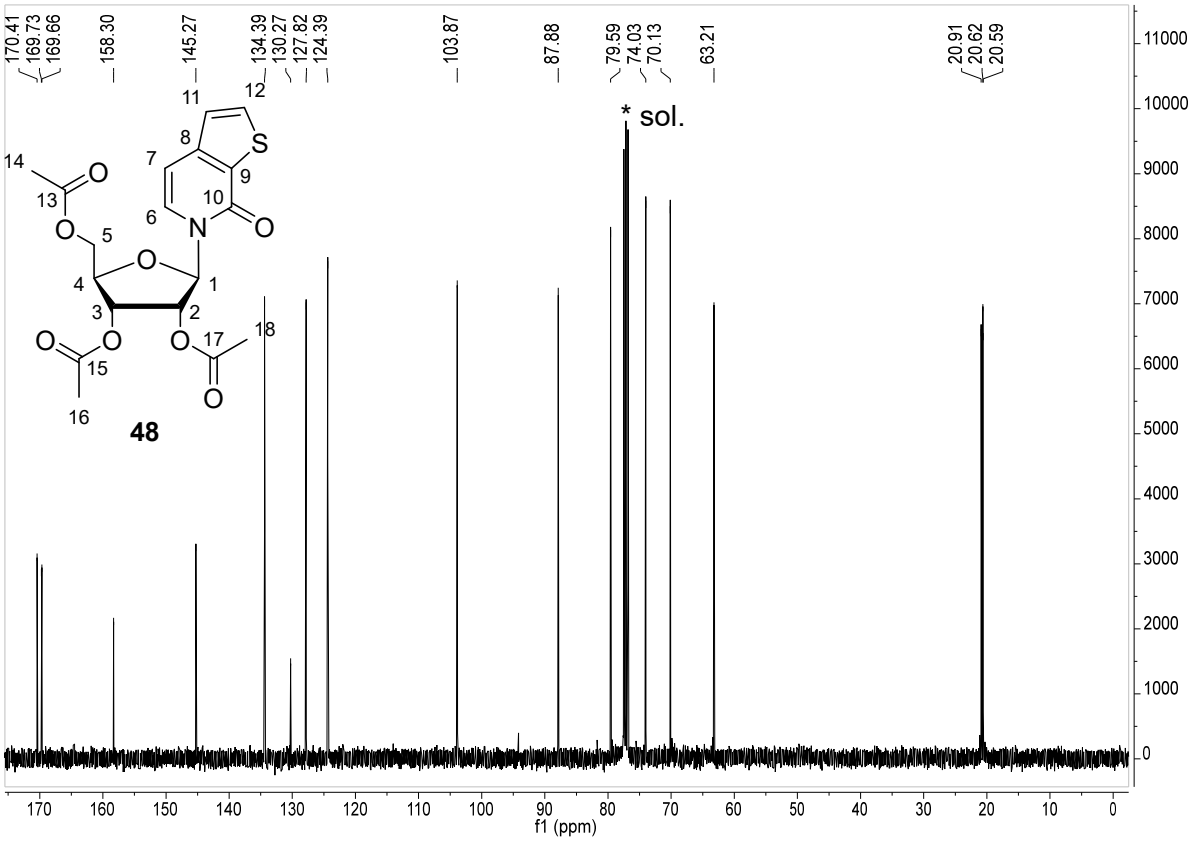


Figure 167. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of compound **48**.

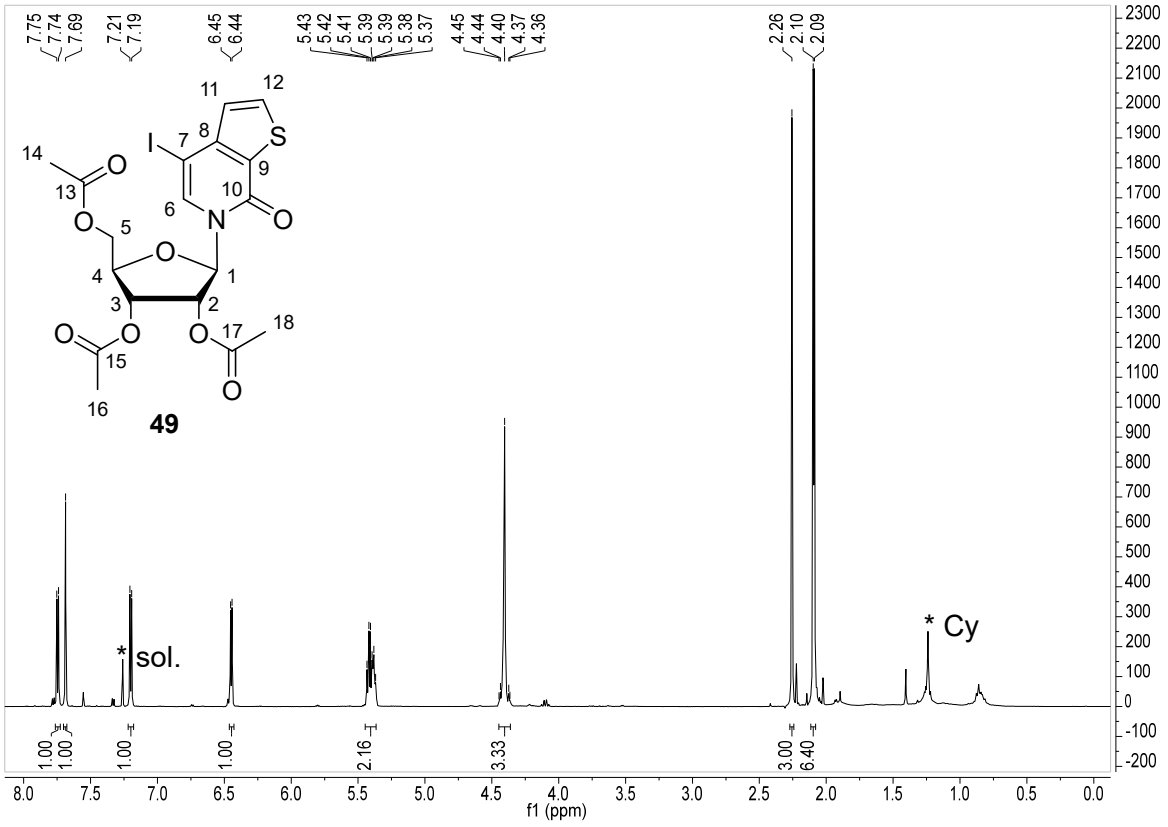


Figure 168. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **49**.

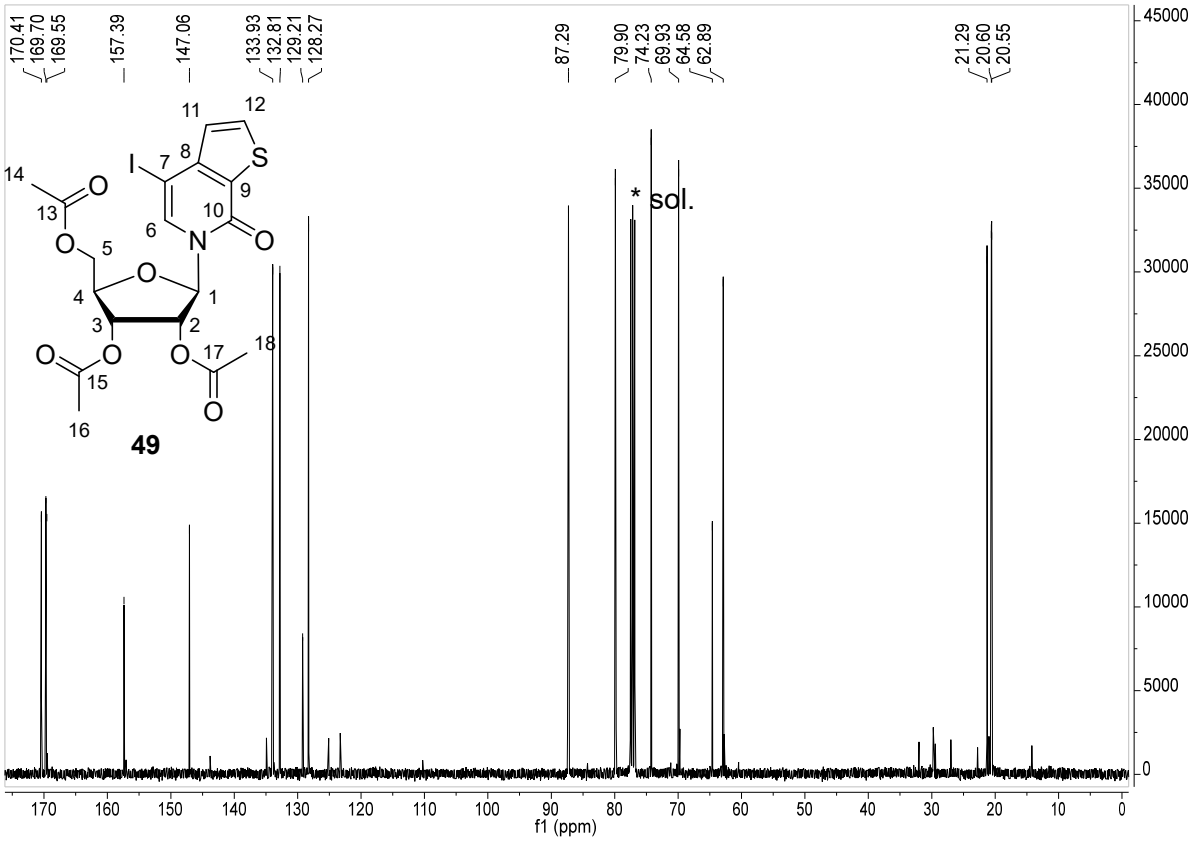


Figure 169. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of compound **49**.

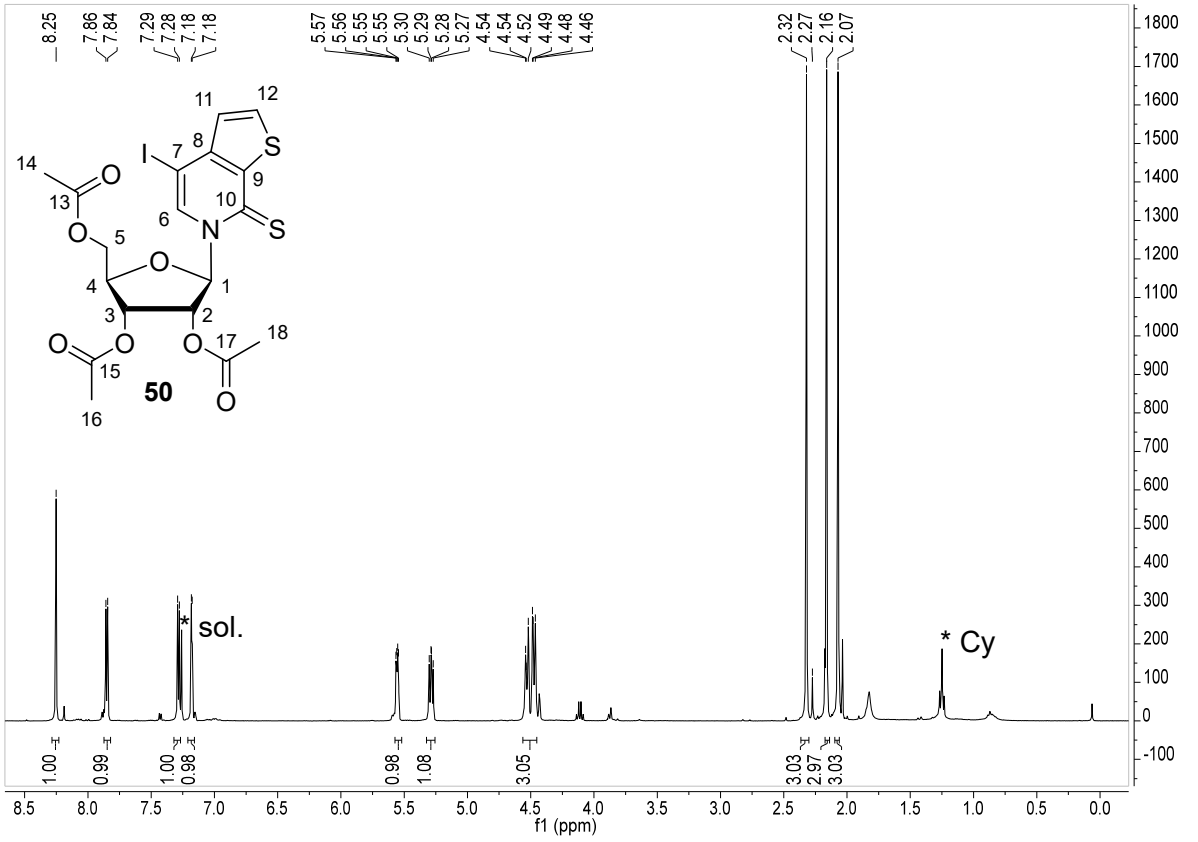


Figure 170. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of thienated compound **50**.

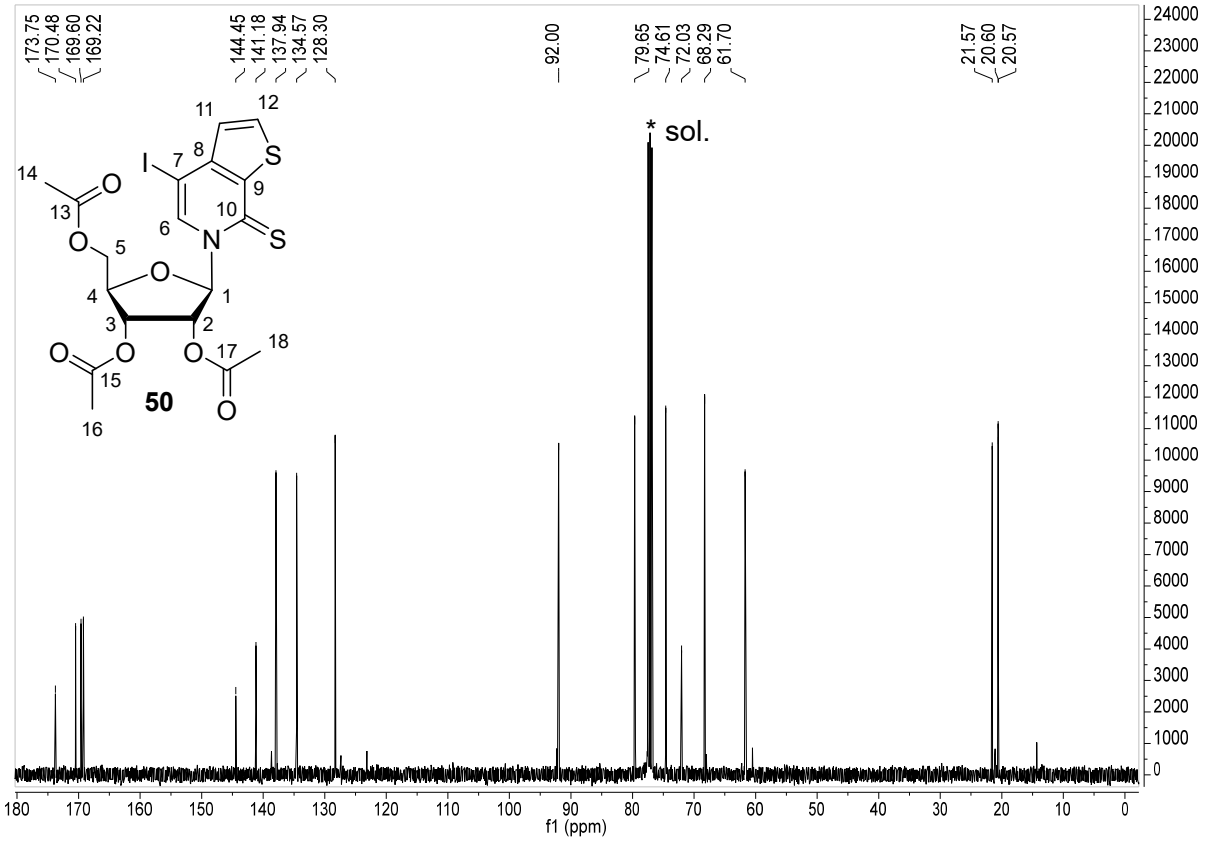


Figure 171. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of thienated compound **50**.

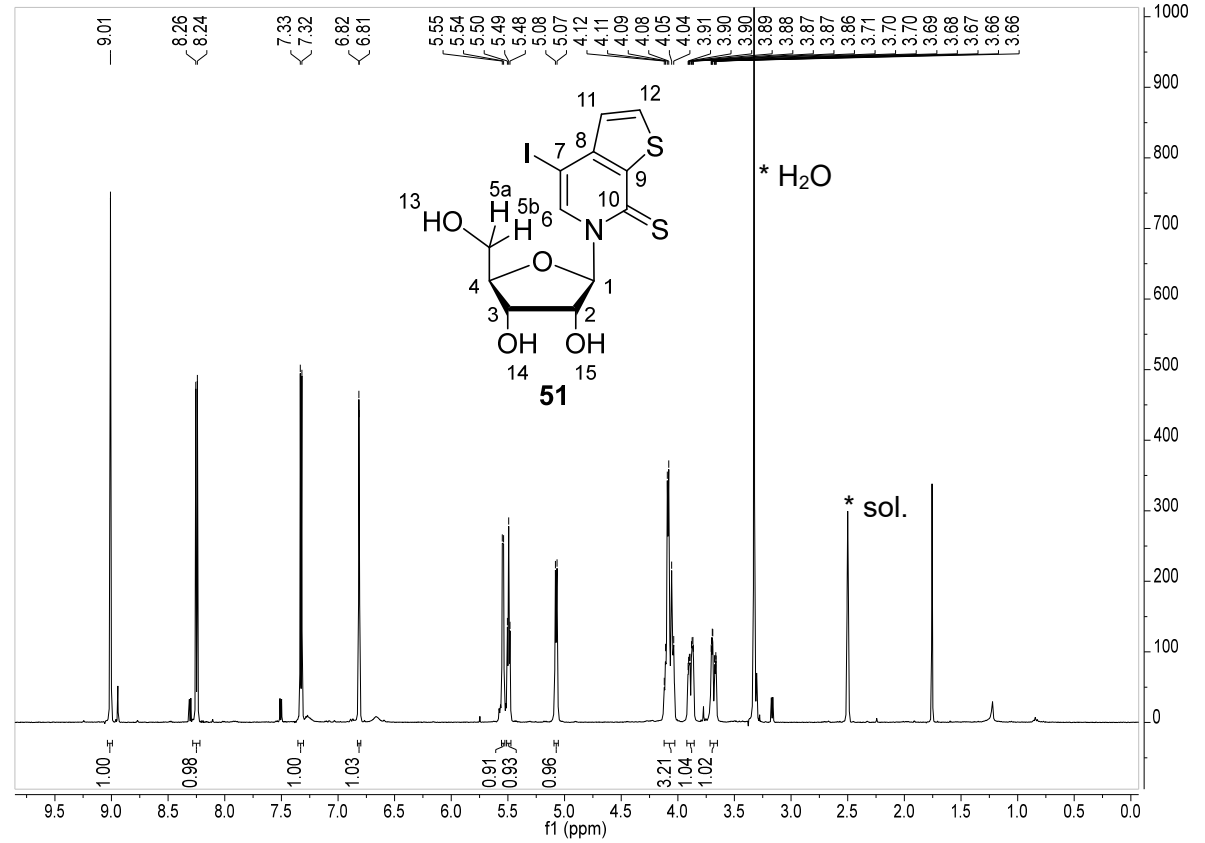


Figure 172. ^{1}H-NMR spectrum (DMSO-d_6, 400 MHz) of nucleoside analog **51**.

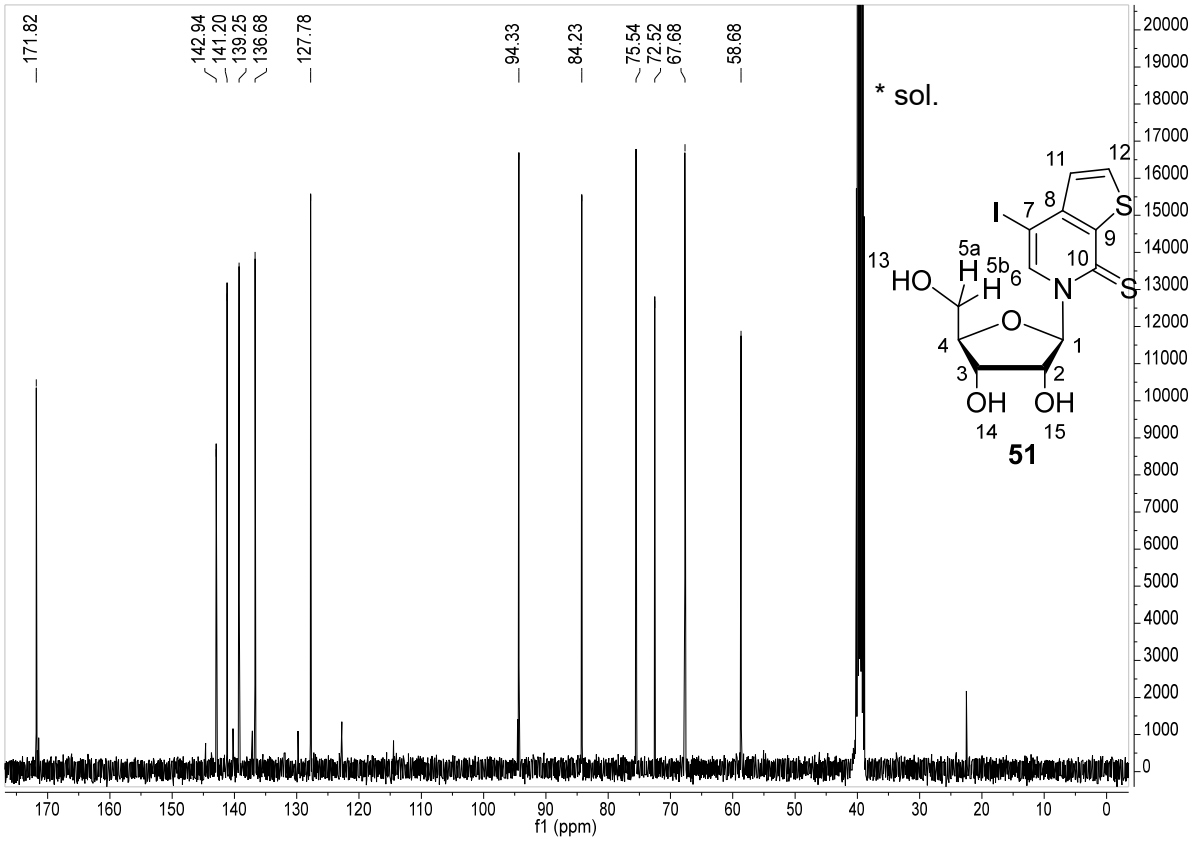


Figure 173. ^{13}C-NMR spectrum (DMSO-d_6, 101 MHz) of nucleoside analog **51**.

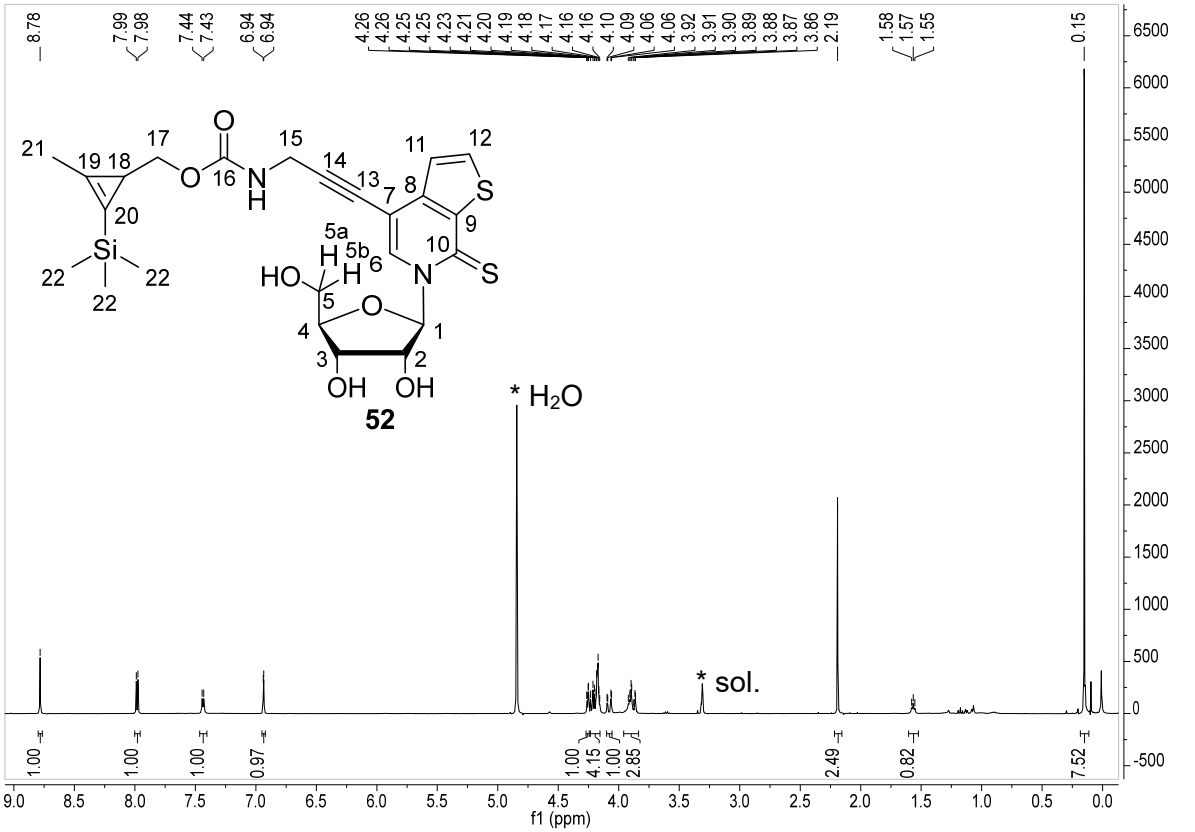


Figure 174. ^{1}H-NMR spectrum (CD_3OD, 400 MHz) of compound **52**.

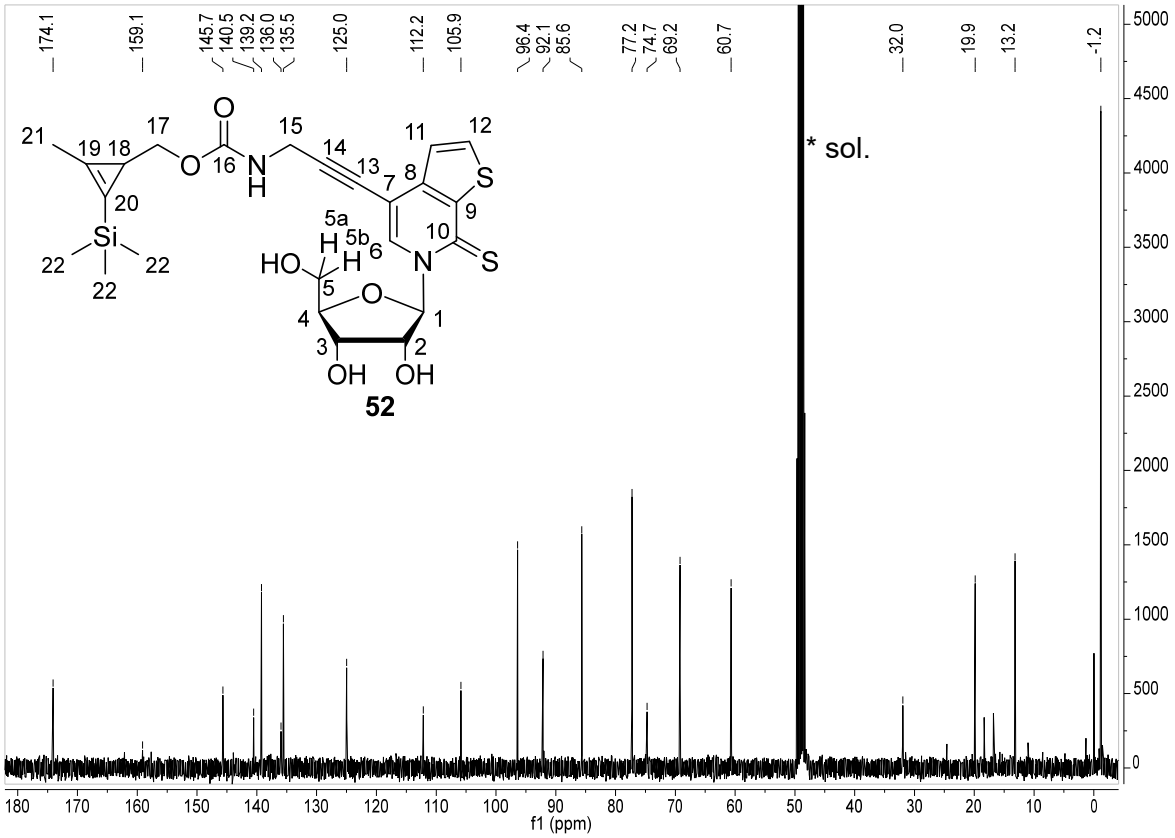


Figure 175. ^{13}C-NMR spectrum (CD_3OD, 101 MHz) of compound **52**.

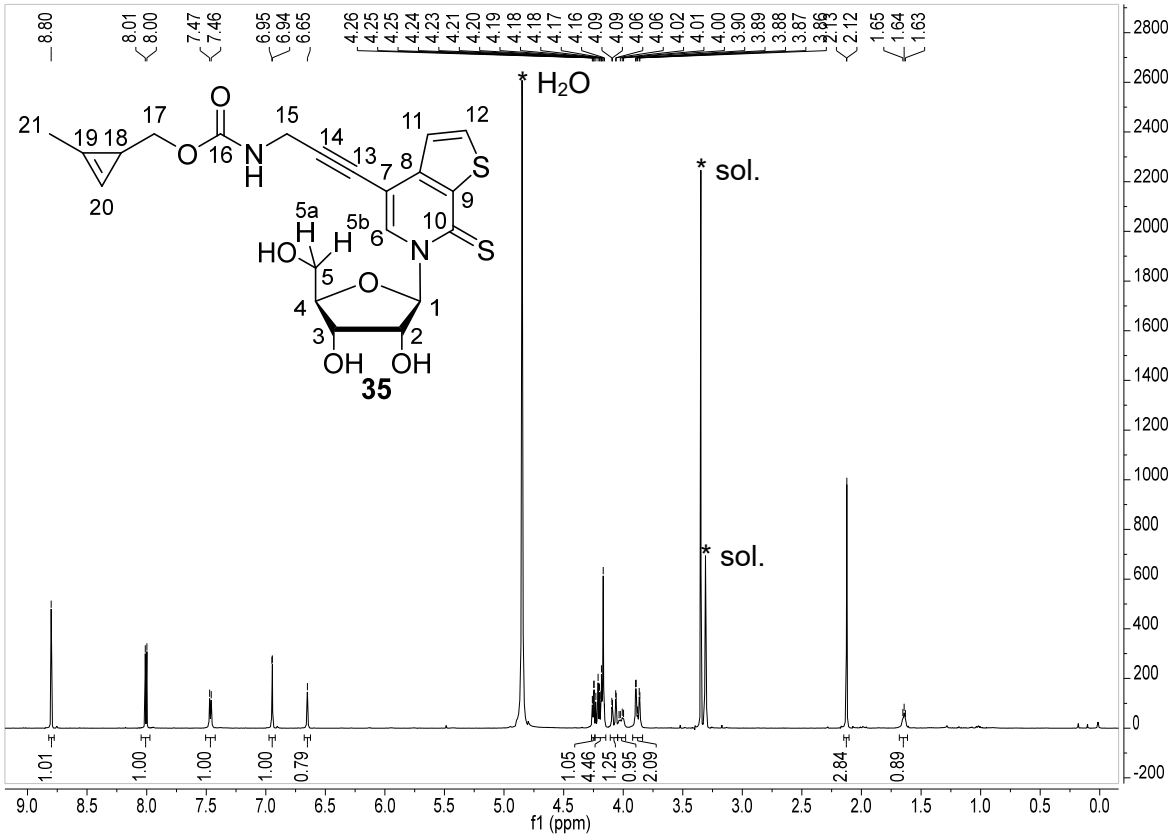


Figure 176. ^{1}H-NMR spectrum (CD_3OD, 400 MHz) of compound **35**.

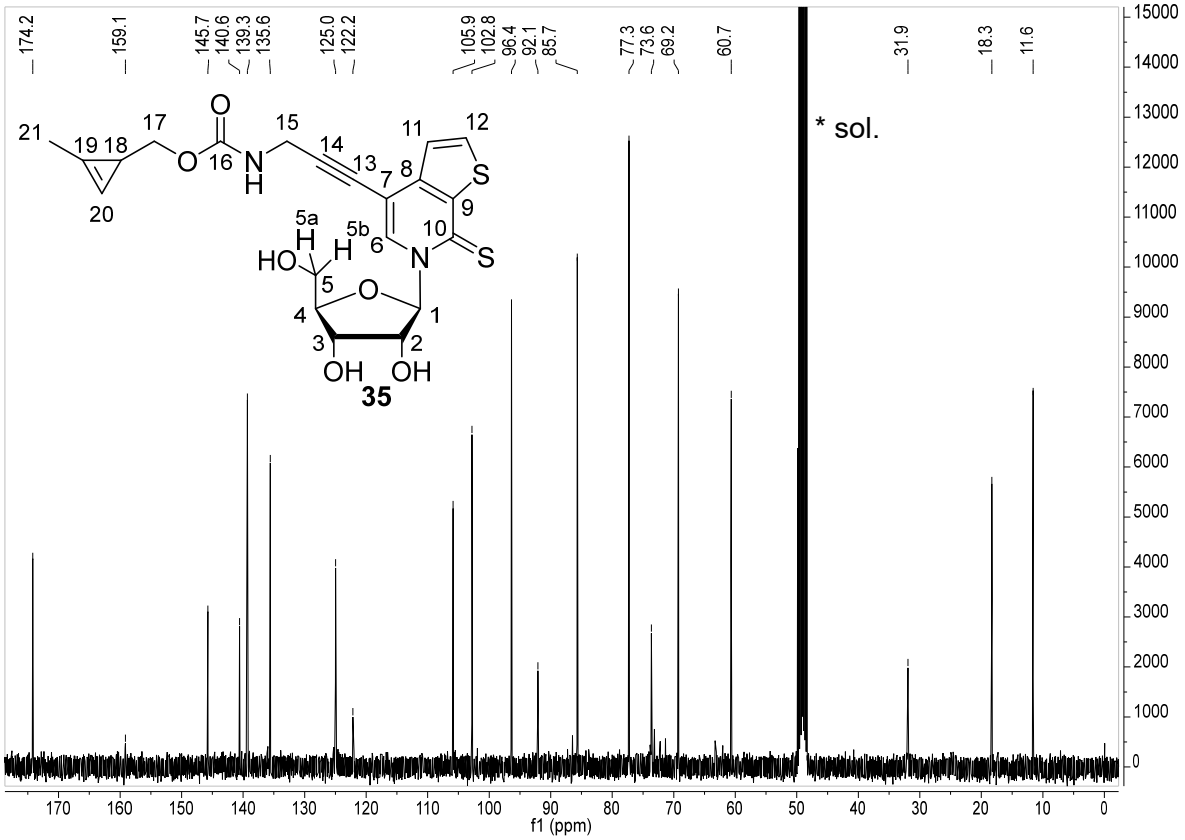


Figure 177. ^{13}C-NMR spectrum (CD_3OD, 101 MHz) of compound **35**.

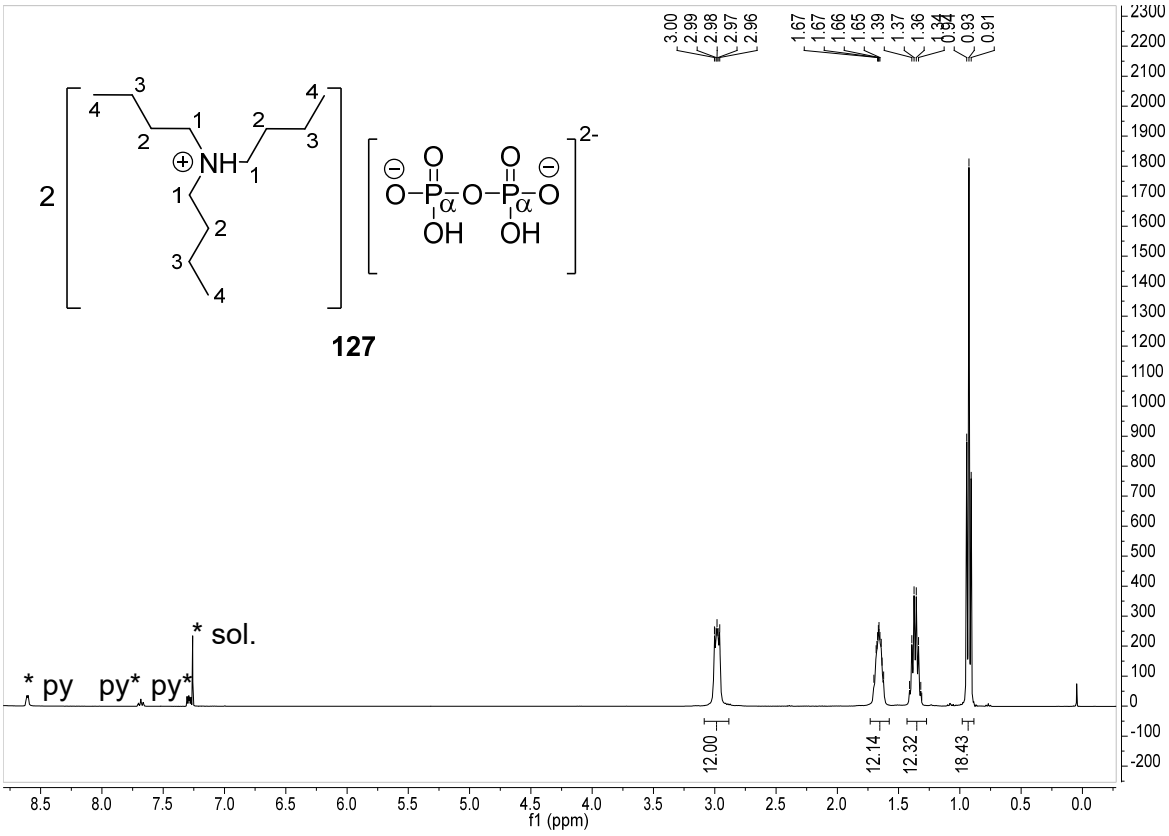
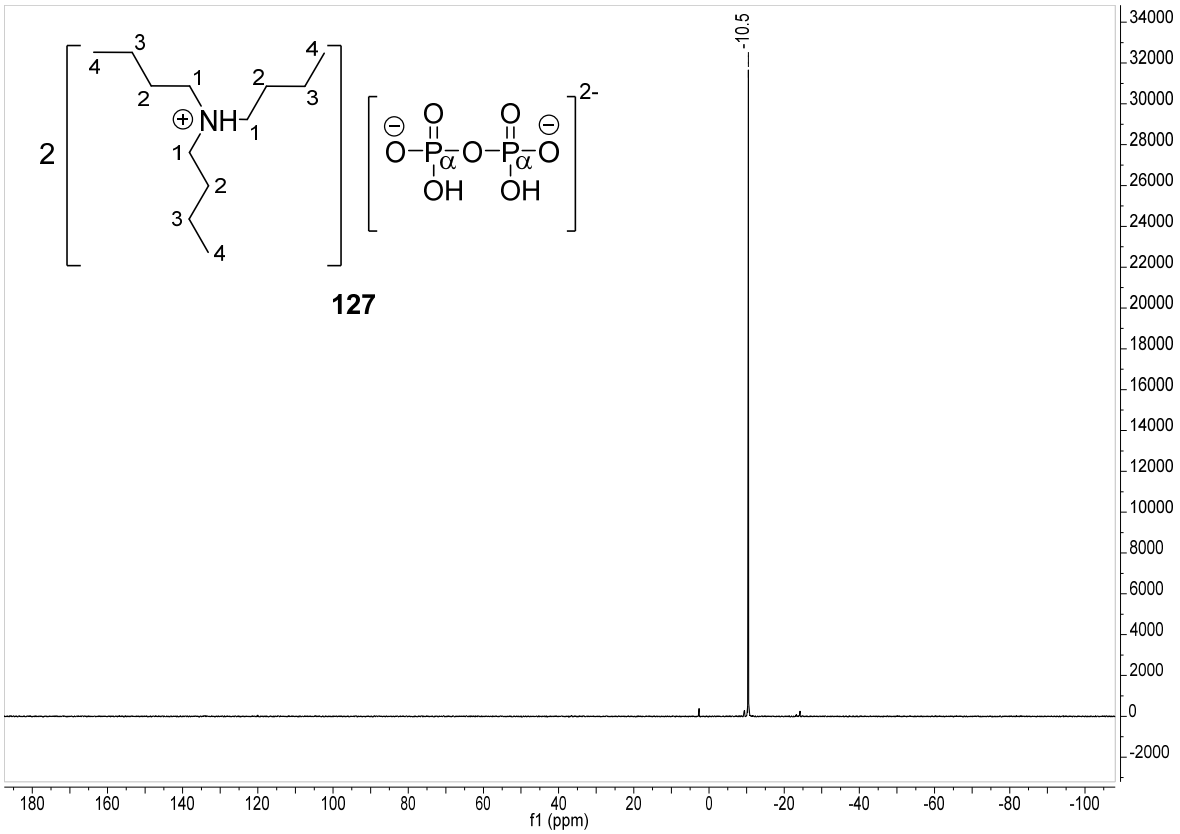
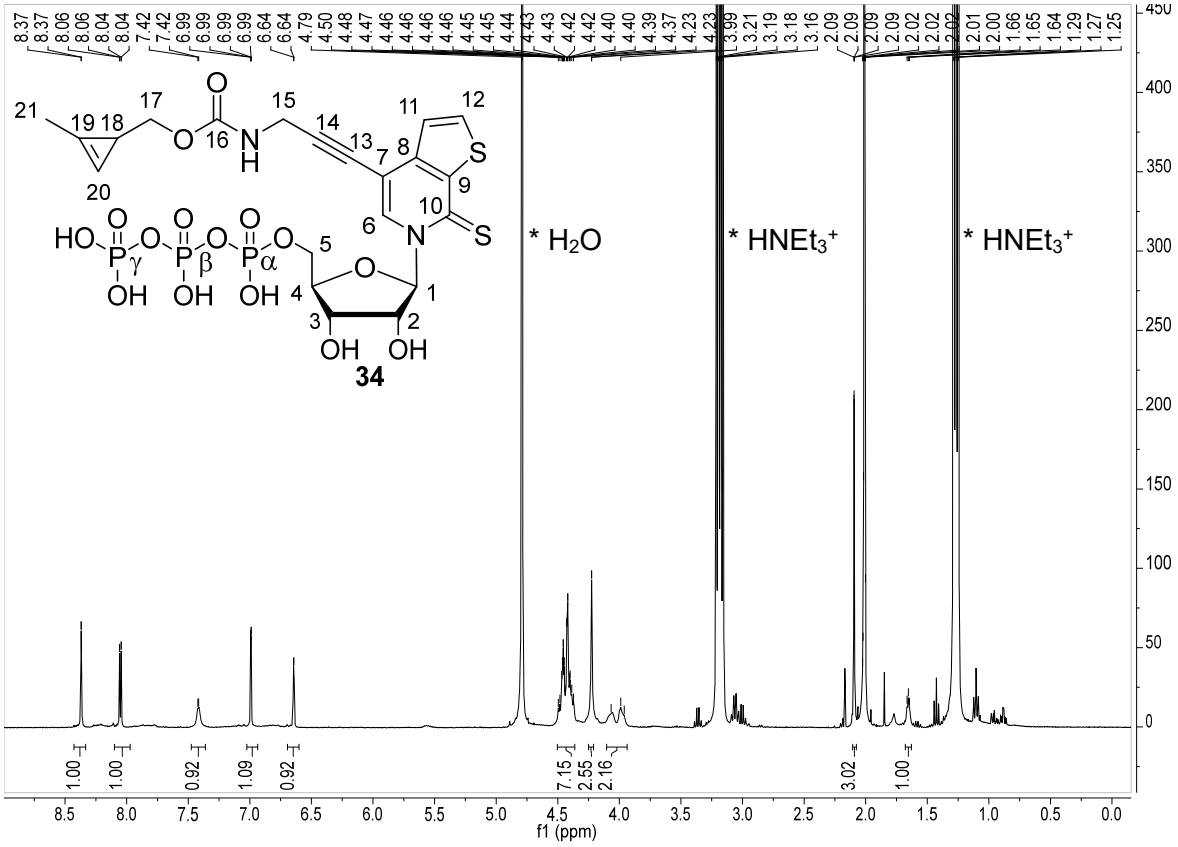


Figure 178. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of **127**.



127

Figure 179. ^{31}P-NMR spectrum ($CDCl_3$, 162 MHz) of **127**.



34

Figure 180. ^{1}H-NMR spectrum (D_2O, 400 MHz) of triphosphate **34**.

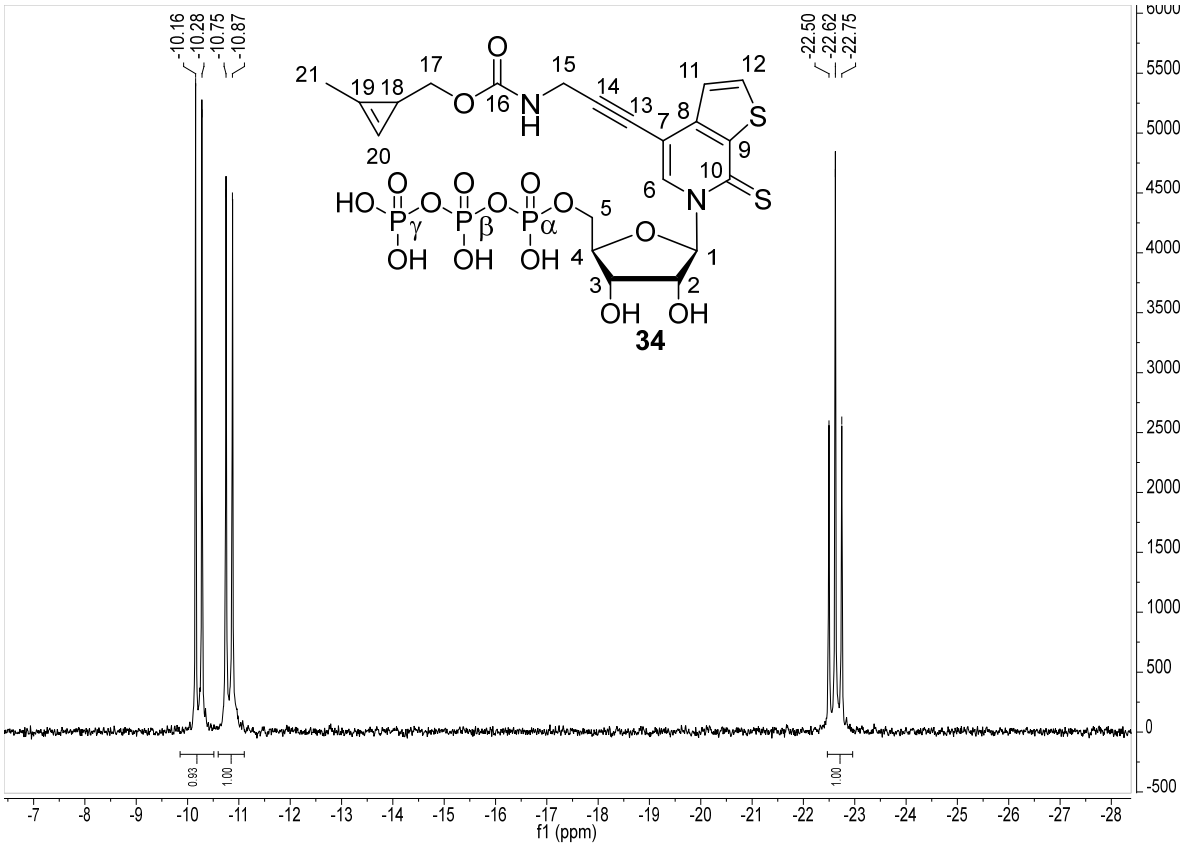


Figure 181. ^{31}P-NMR spectrum (D_2O, 162 MHz) of triphosphate **34**.

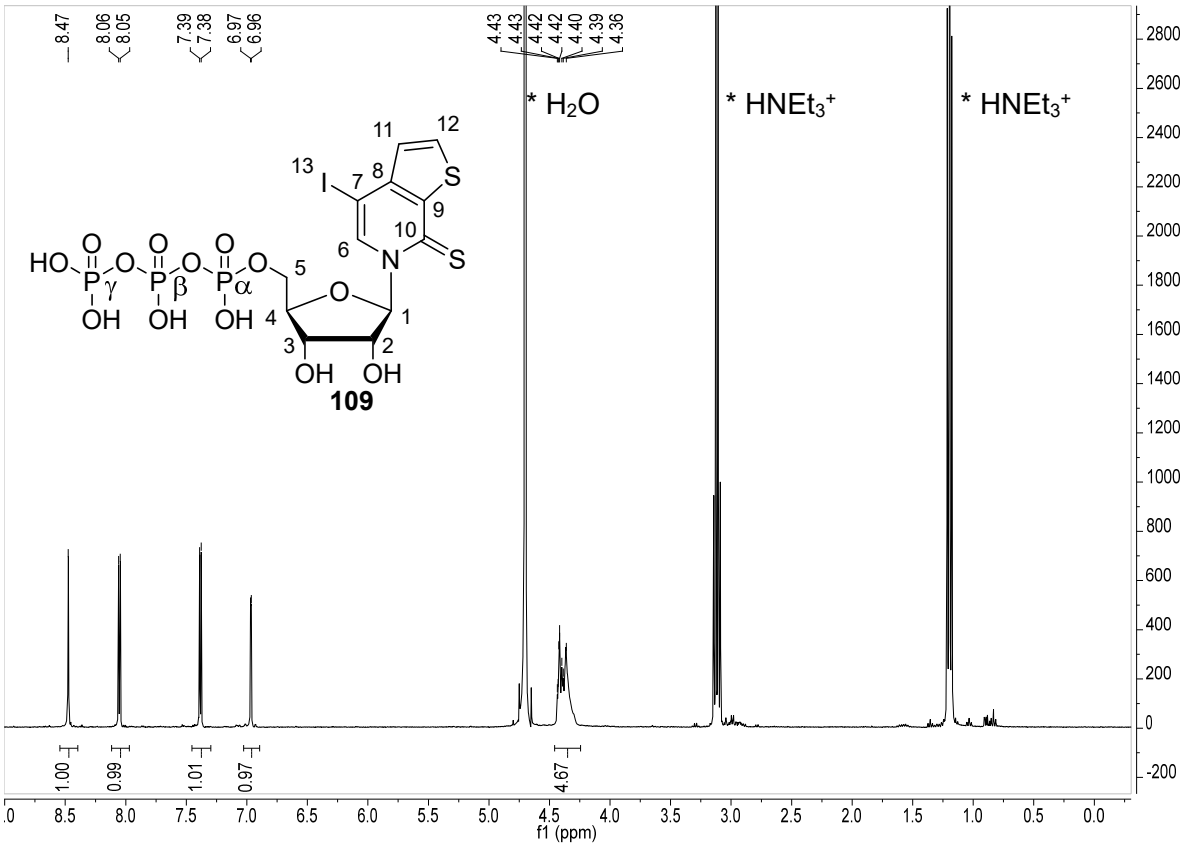


Figure 182. ^{1}H-NMR spectrum (D_2O, 400 MHz) of triphosphate **109**.

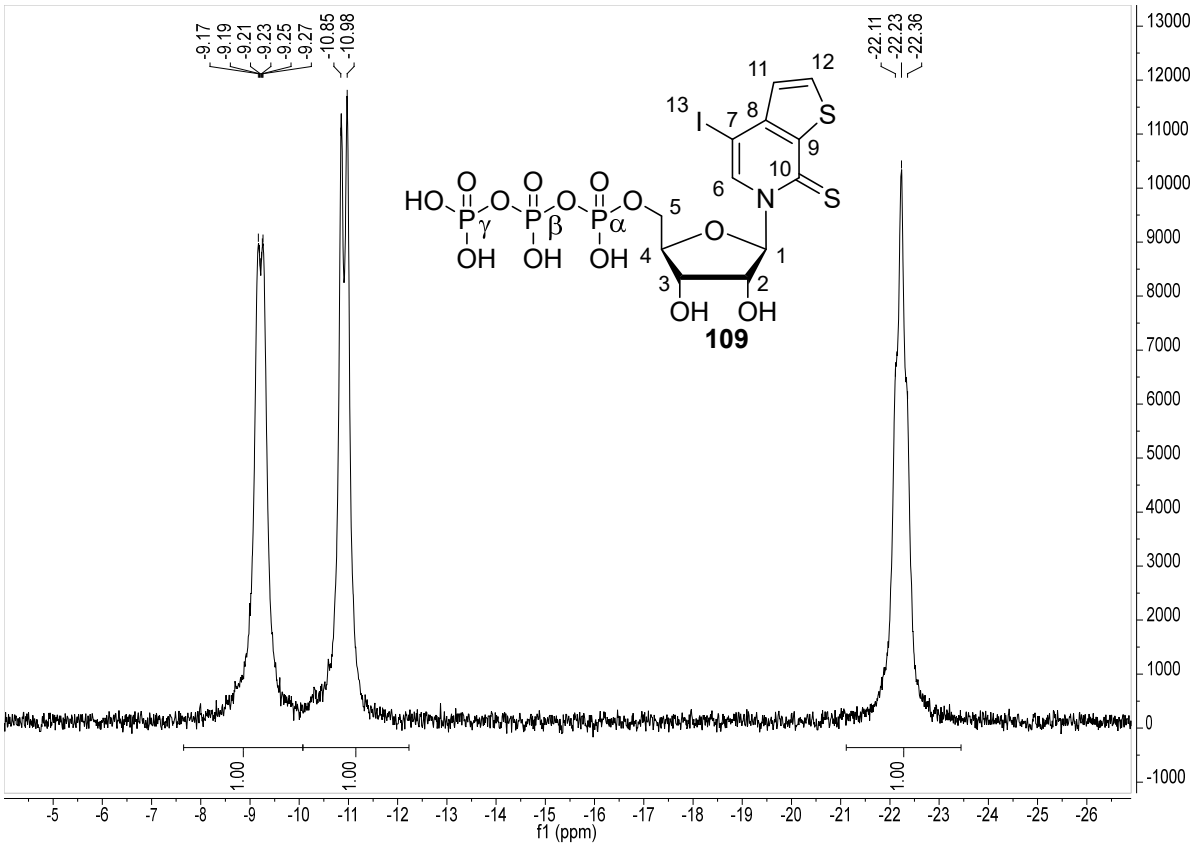


Figure 183. ^{31}P-NMR spectrum (D_2O, 162 MHz) of triphosphate **109**.

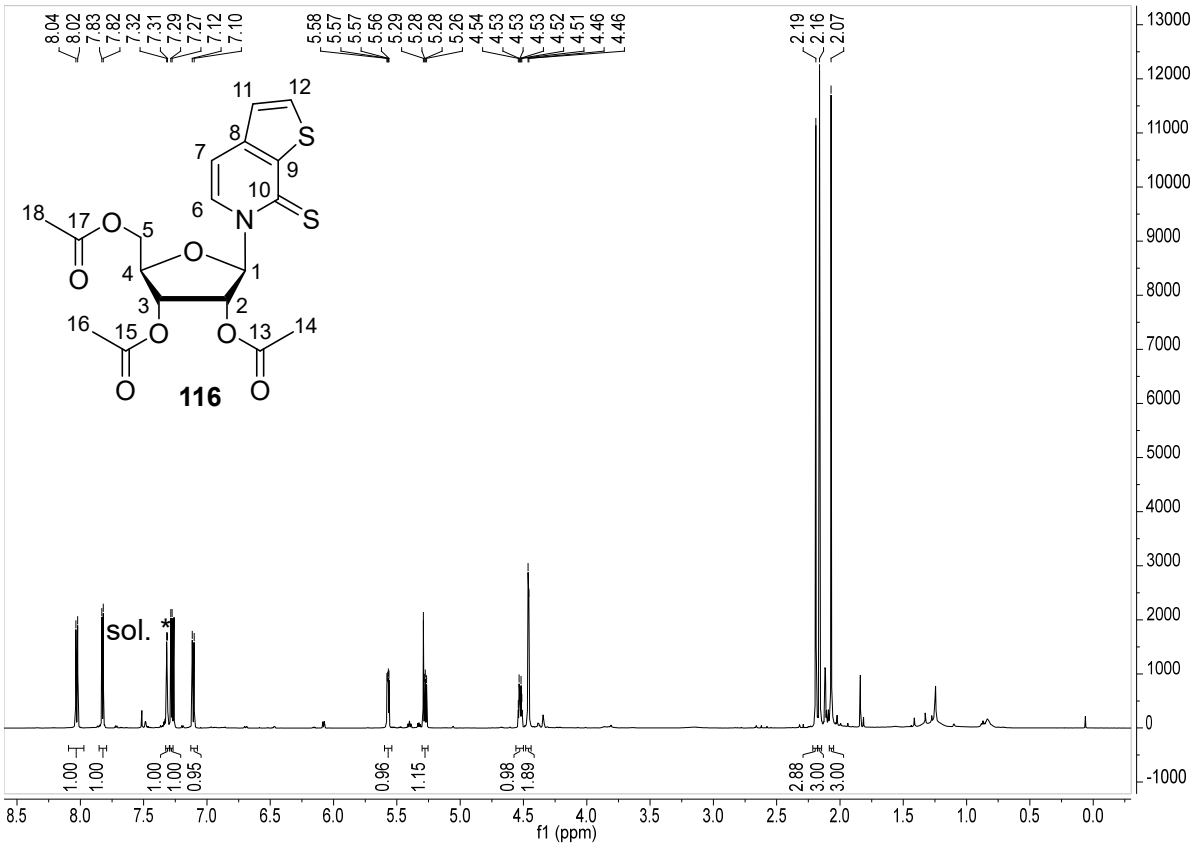


Figure 184. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **116**.

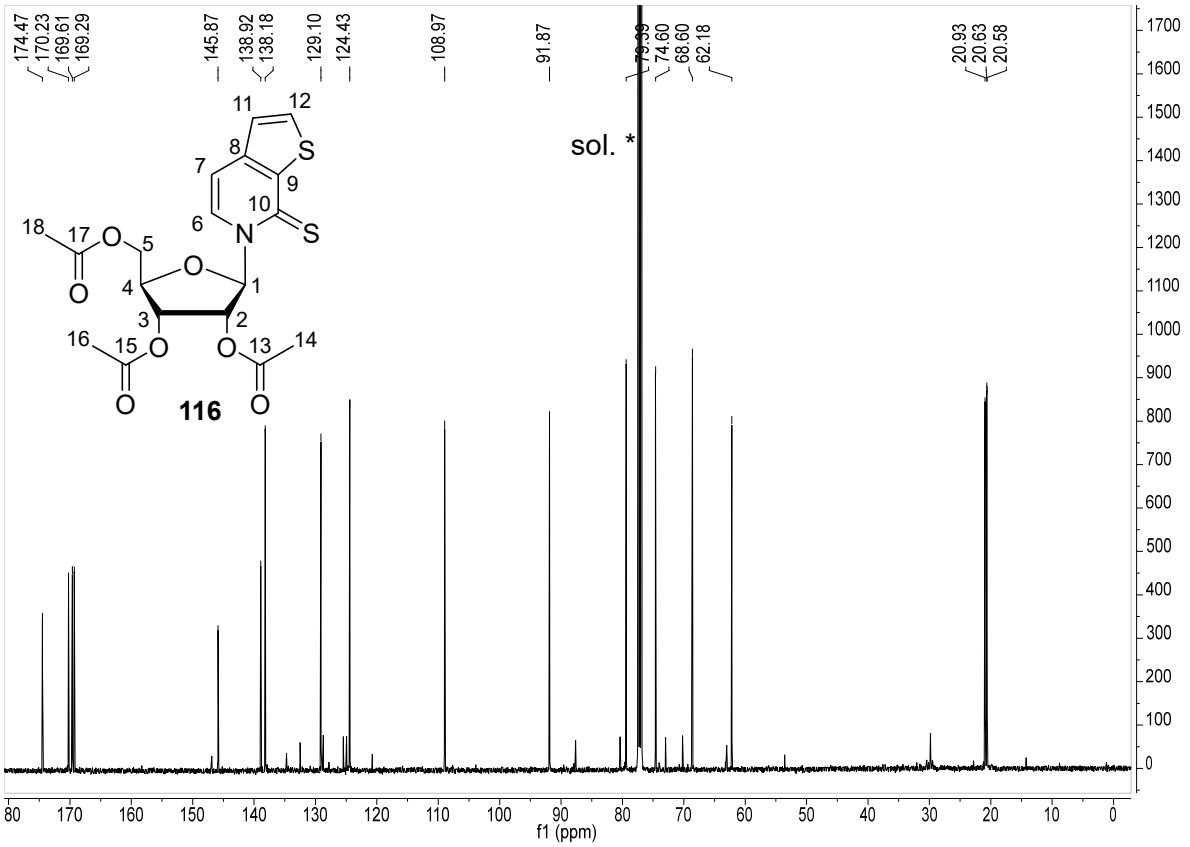


Figure 185. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of compound **116**.

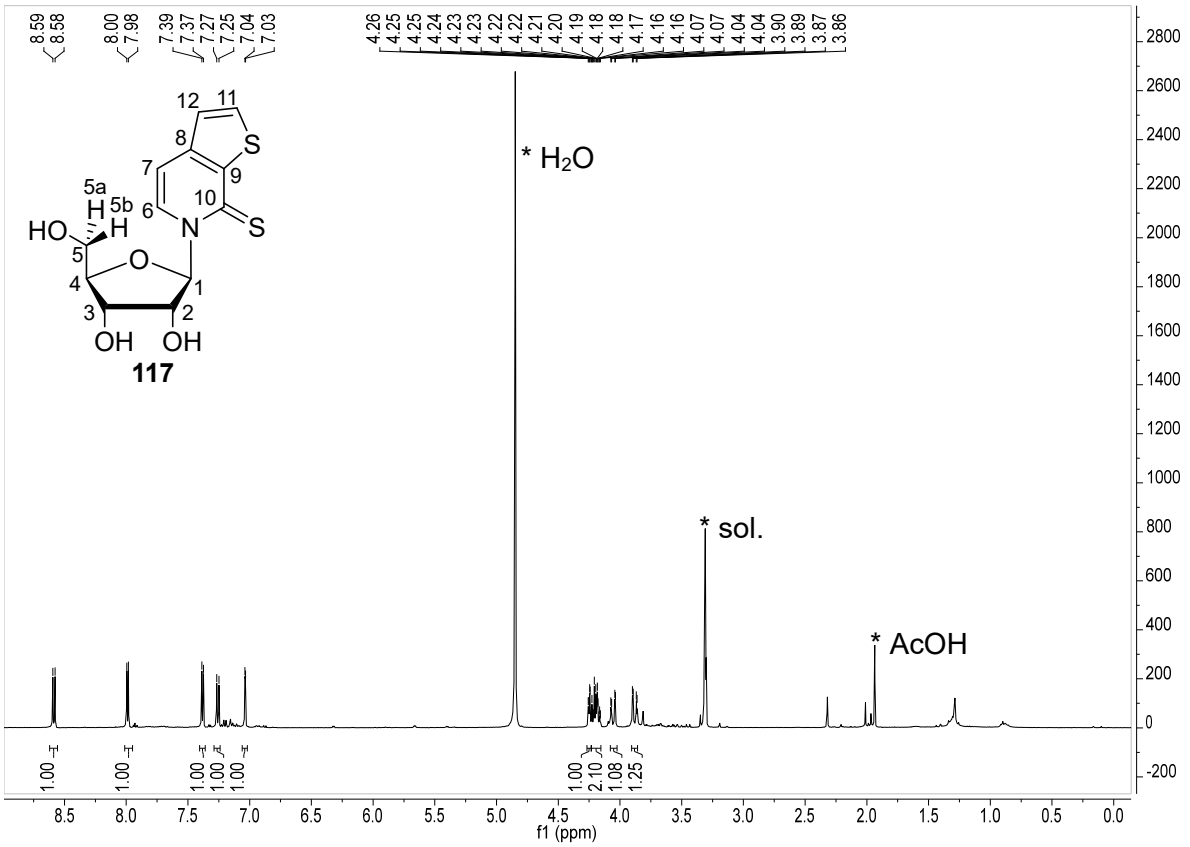


Figure 186. ^{1}H-NMR spectrum (CD_3OD, 400 MHz) of nucleoside analog **117**.

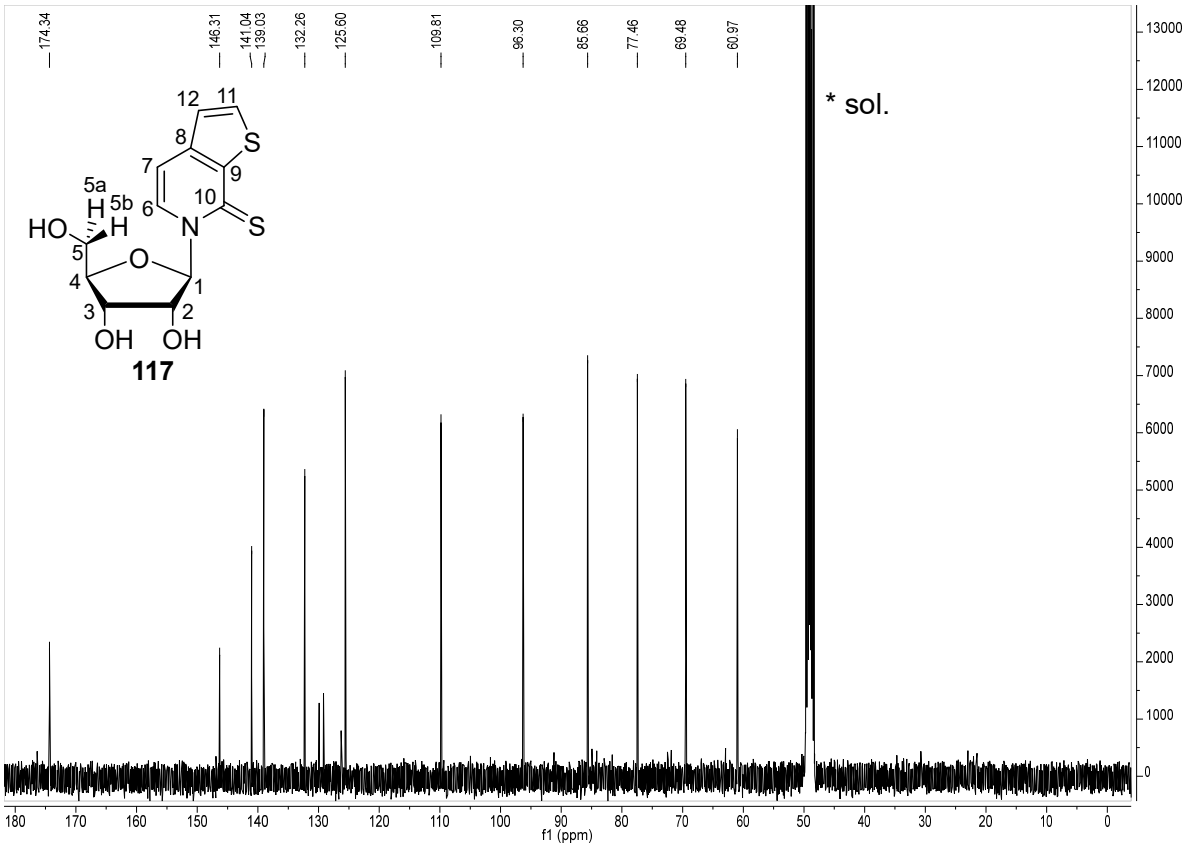


Figure 187. ^{13}C-NMR spectrum (CD_3OD, 101 MHz) of nucleoside analog **117**.

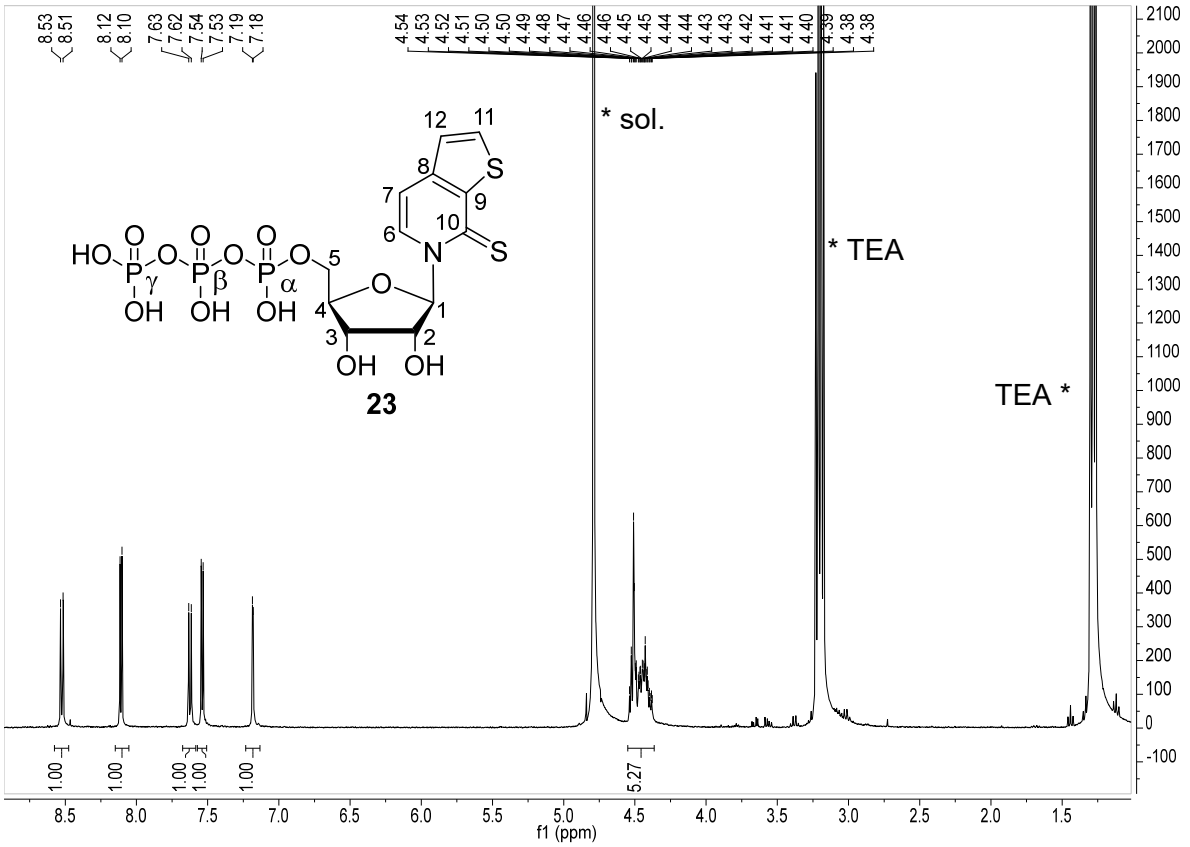


Figure 188. ^{1}H-NMR (D_2O, 400 MHz) of triphosphate **23**.

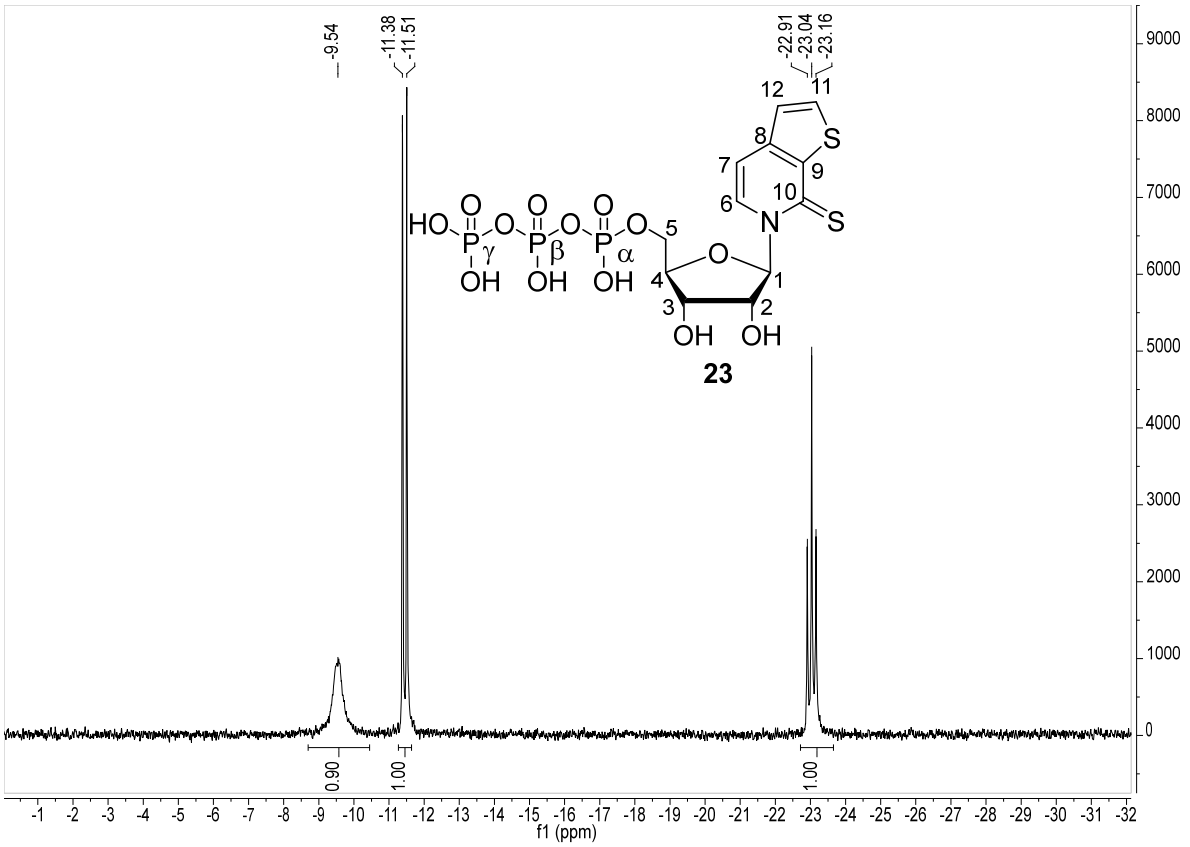


Figure 189. ^{31}P-NMR (D_2O, 162 MHz) of triphosphate **23**.

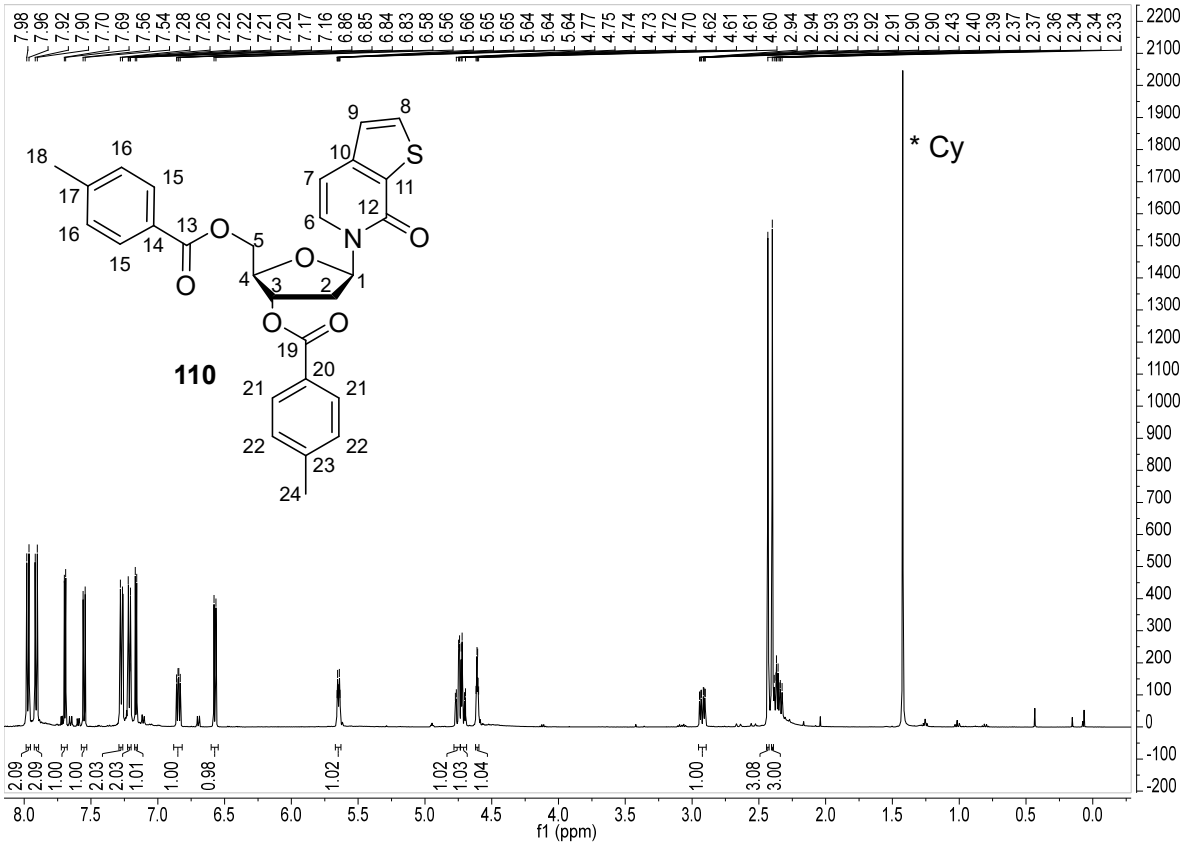


Figure 190. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **110**.

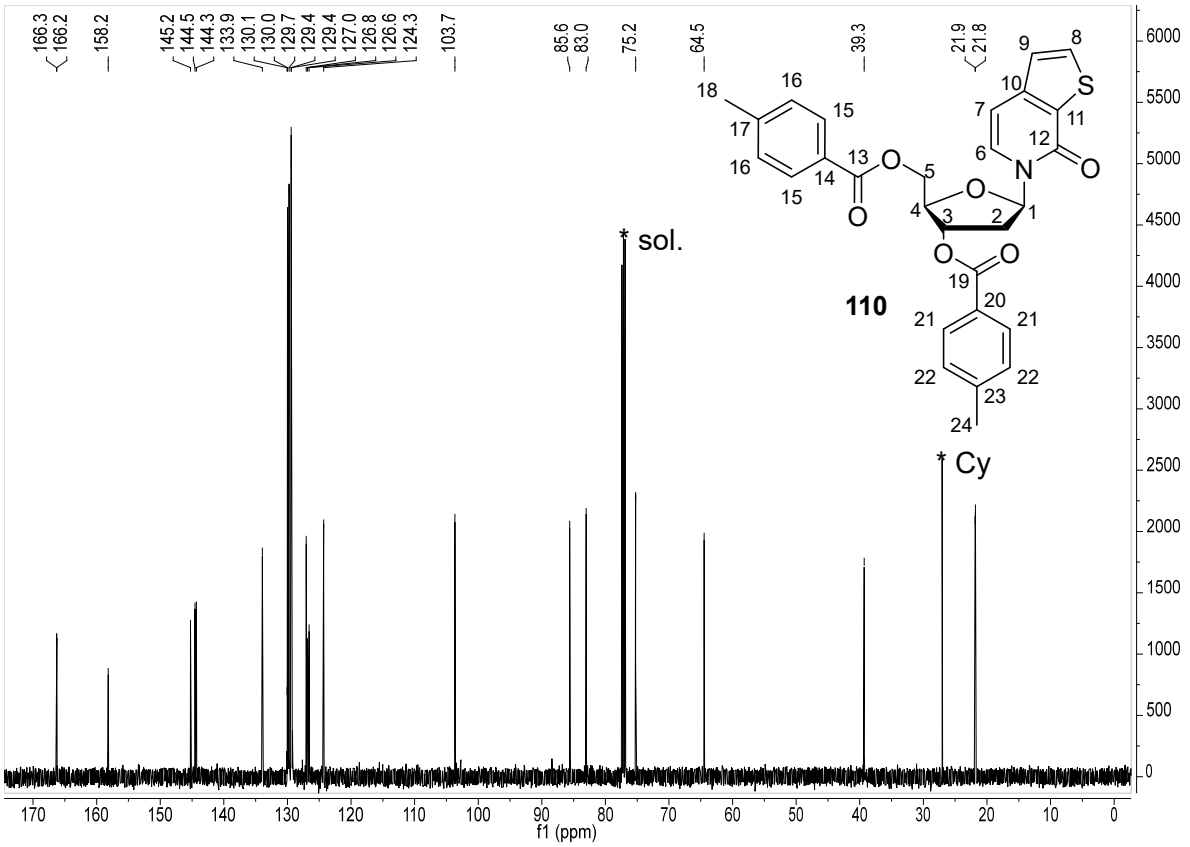


Figure 191. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of compound **110**.

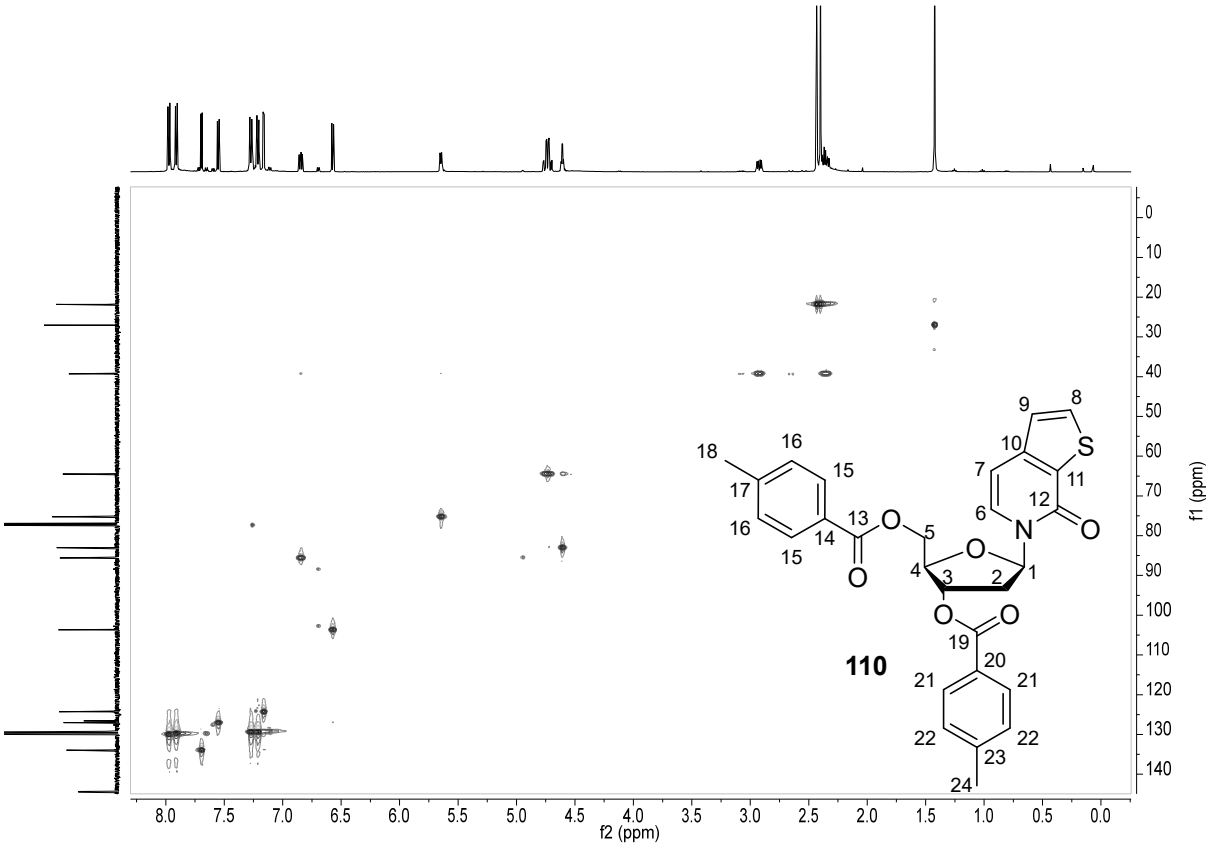


Figure 192. HSQC-NMR spectrum of compound **110**.

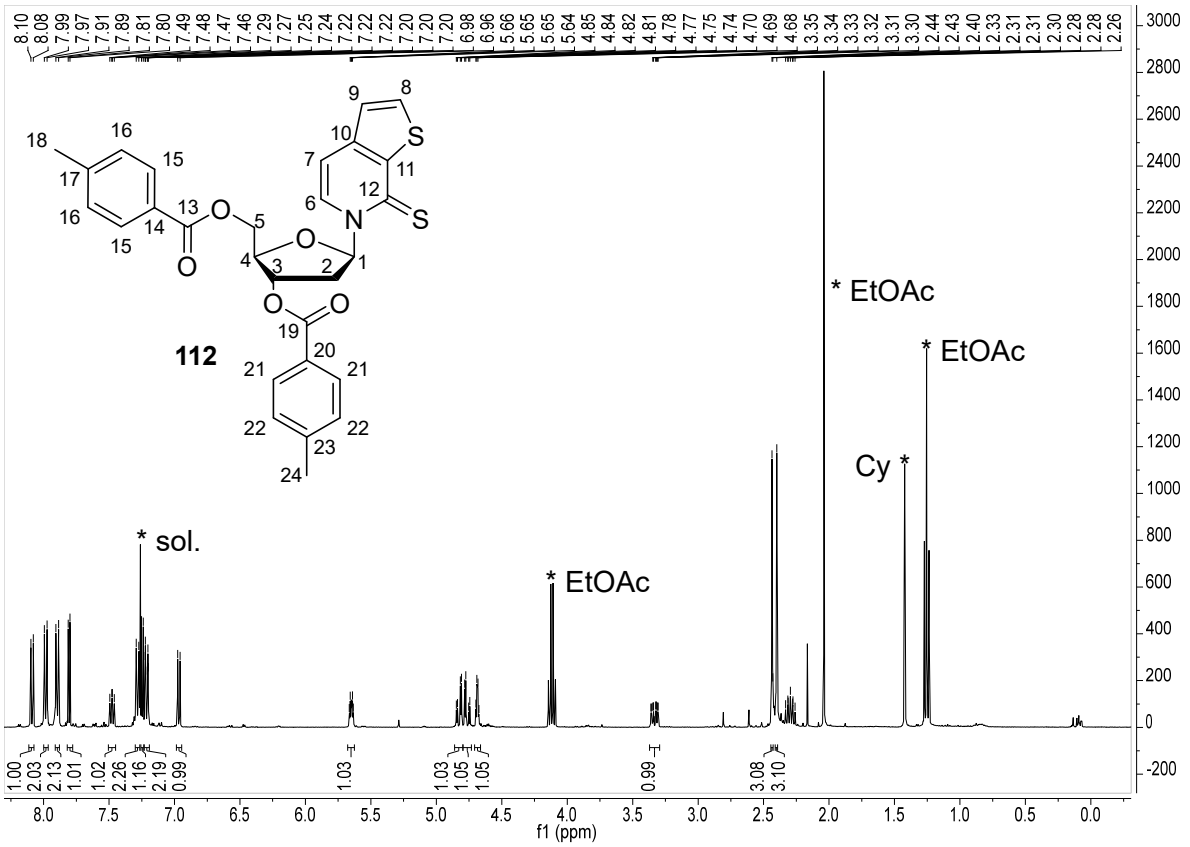


Figure 193. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **112**.

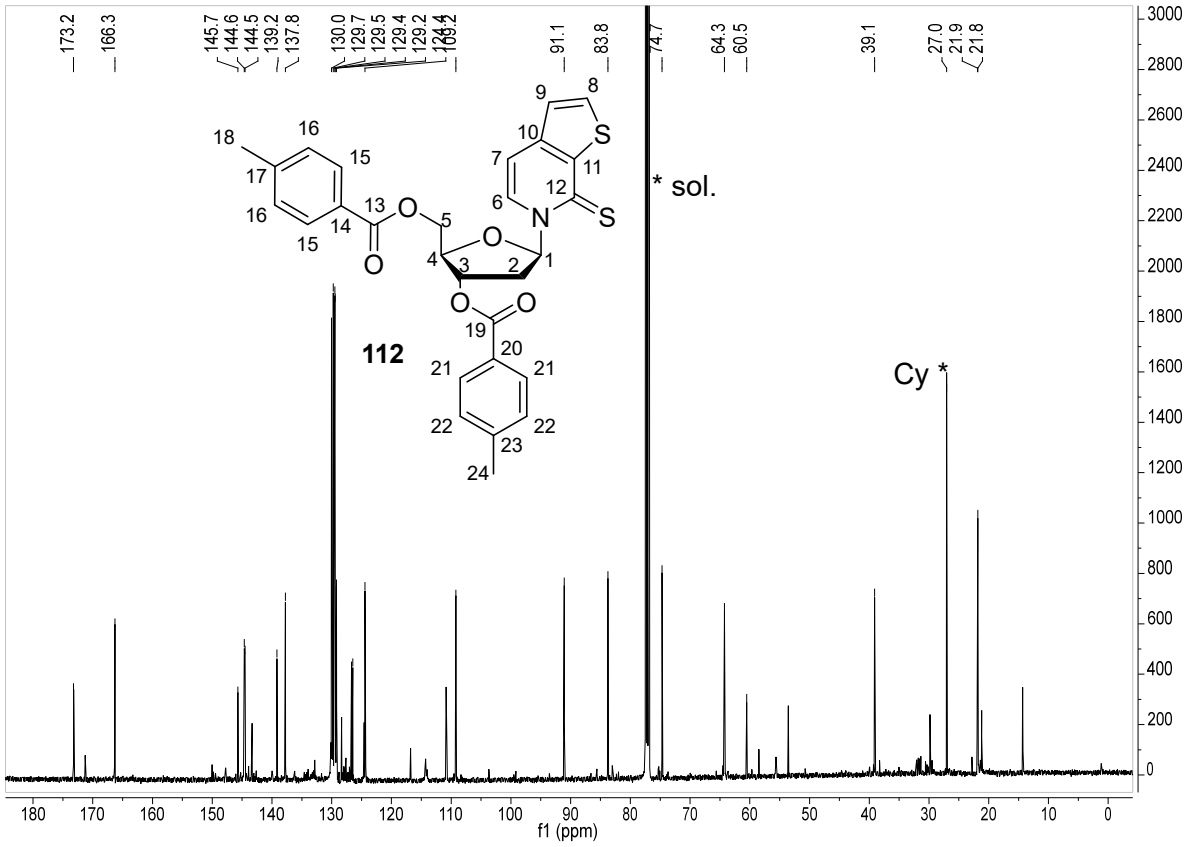


Figure 194. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of compound **112**.

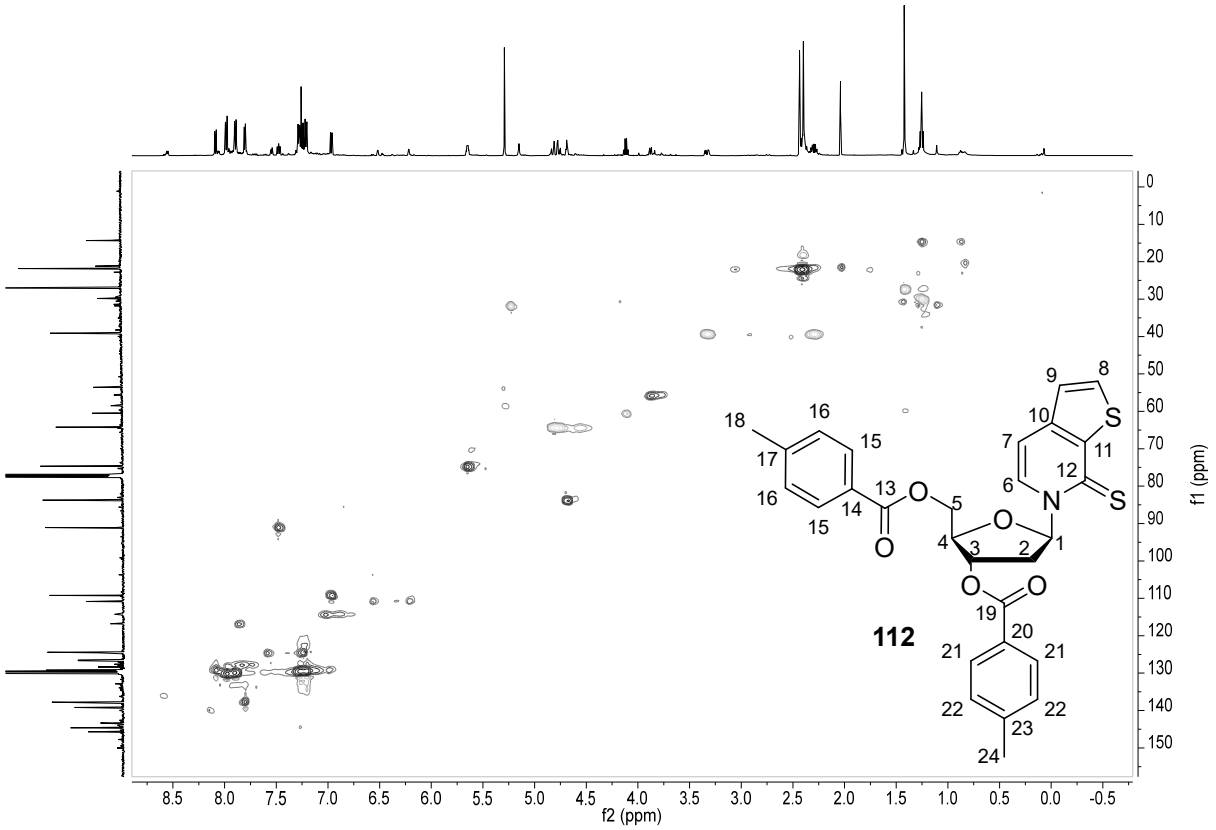


Figure 195. HSQC-NMR spectrum ($CDCl_3$) of compound **112**.

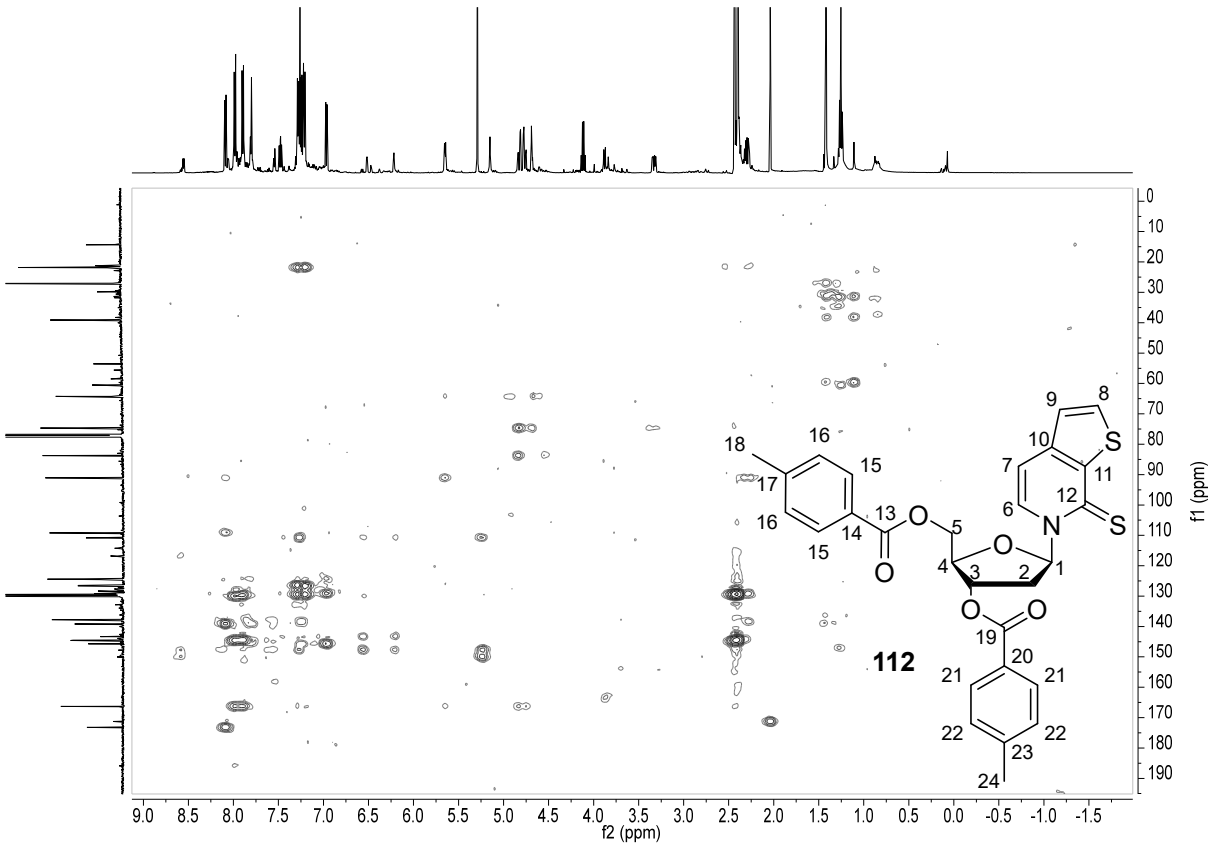


Figure 196. HMBC-NMR spectrum ($CDCl_3$) of compound **112**.

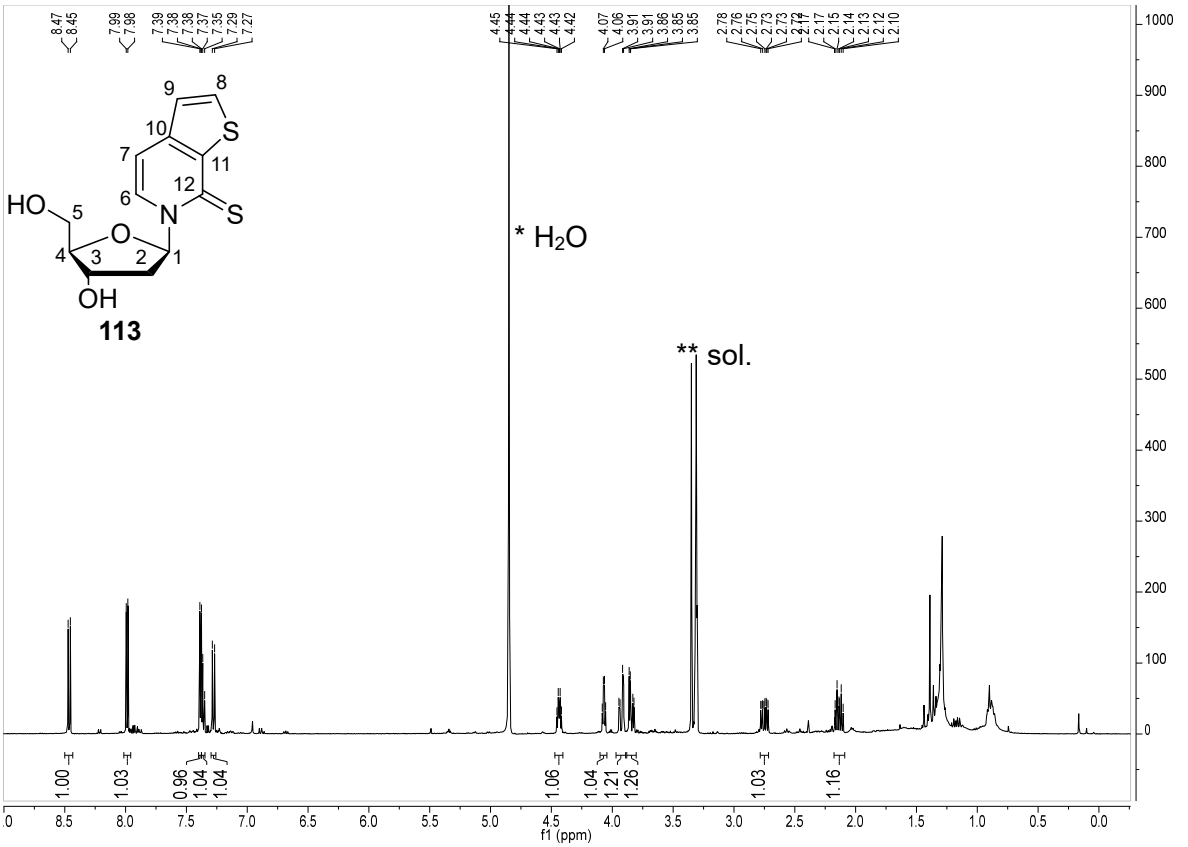


Figure 197. ^{1}H-NMR spectrum (CD_3OD, 400 MHz) of compound **113**.

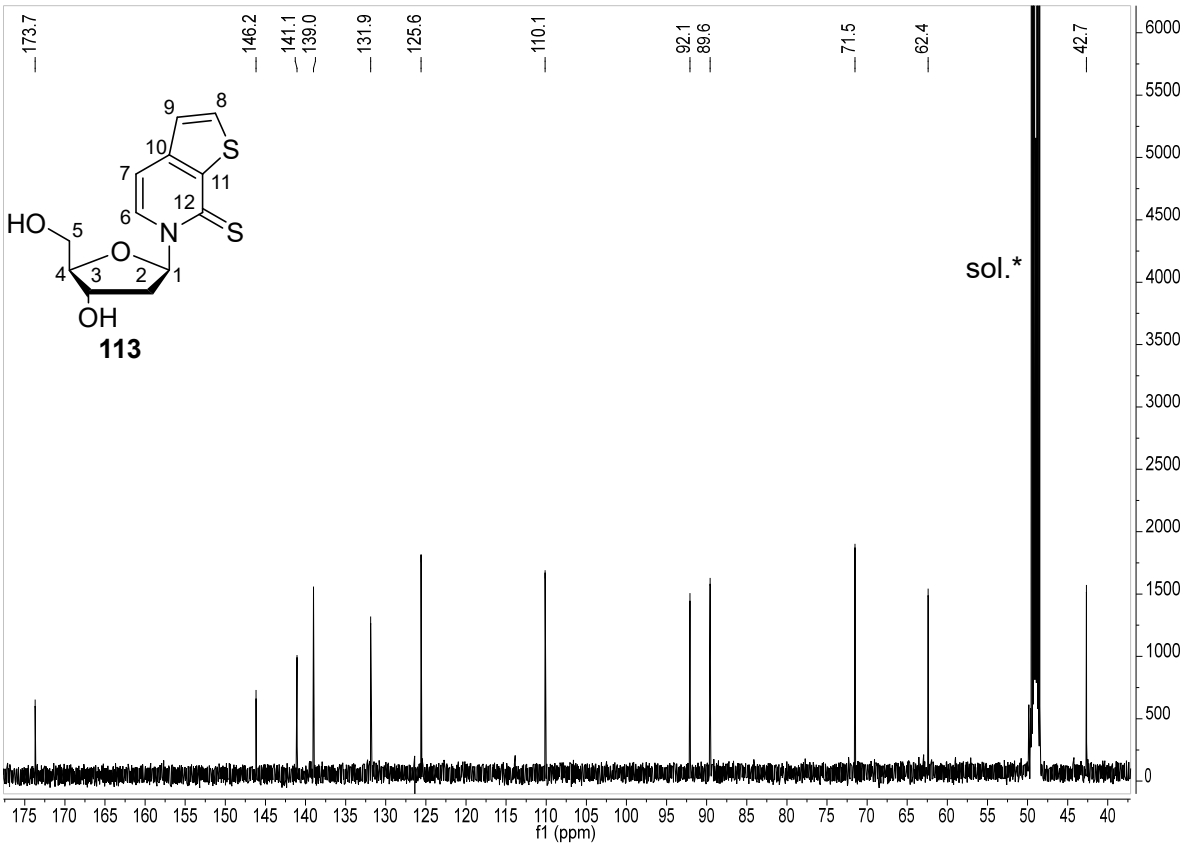


Figure 198. ^{13}C-NMR spectrum (CD_3OD, 101 MHz) of compound **113**.

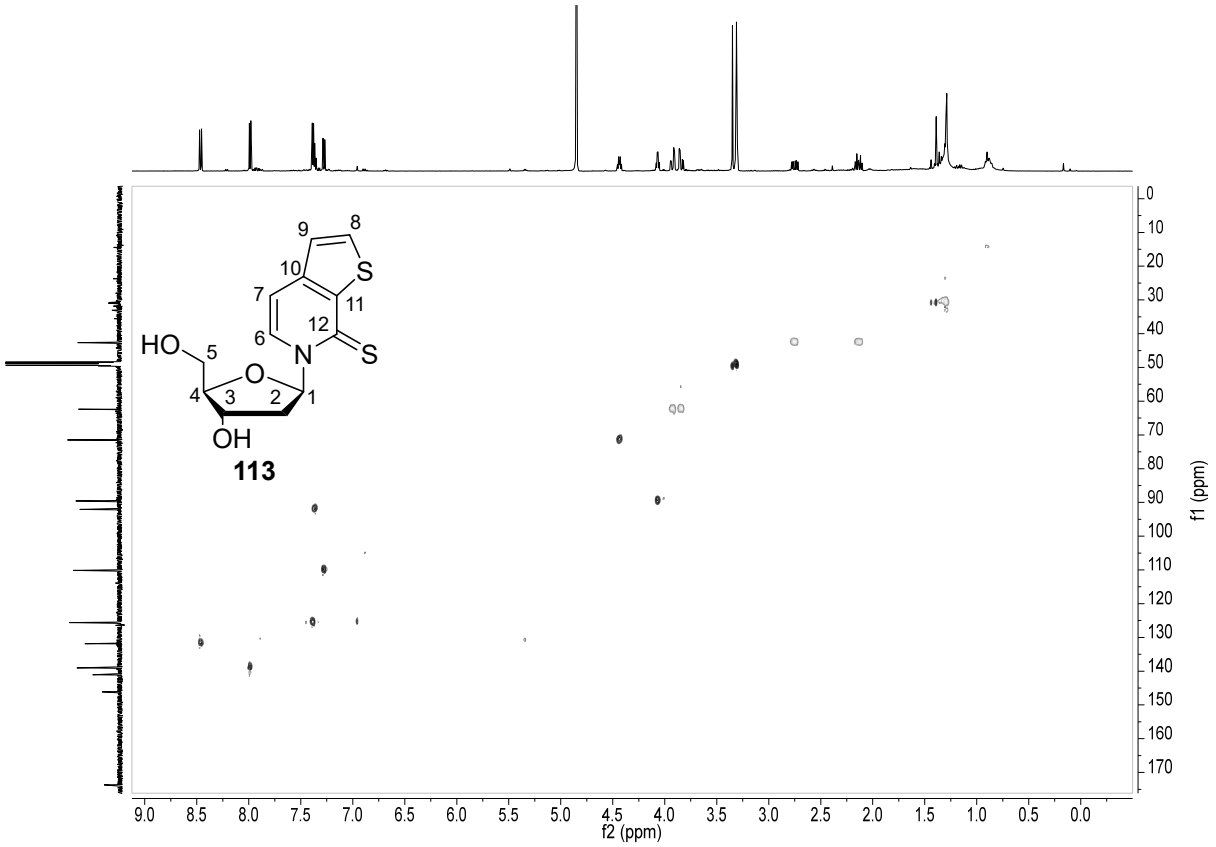


Figure 199. HSQC-NMR spectrum of compound **113**.

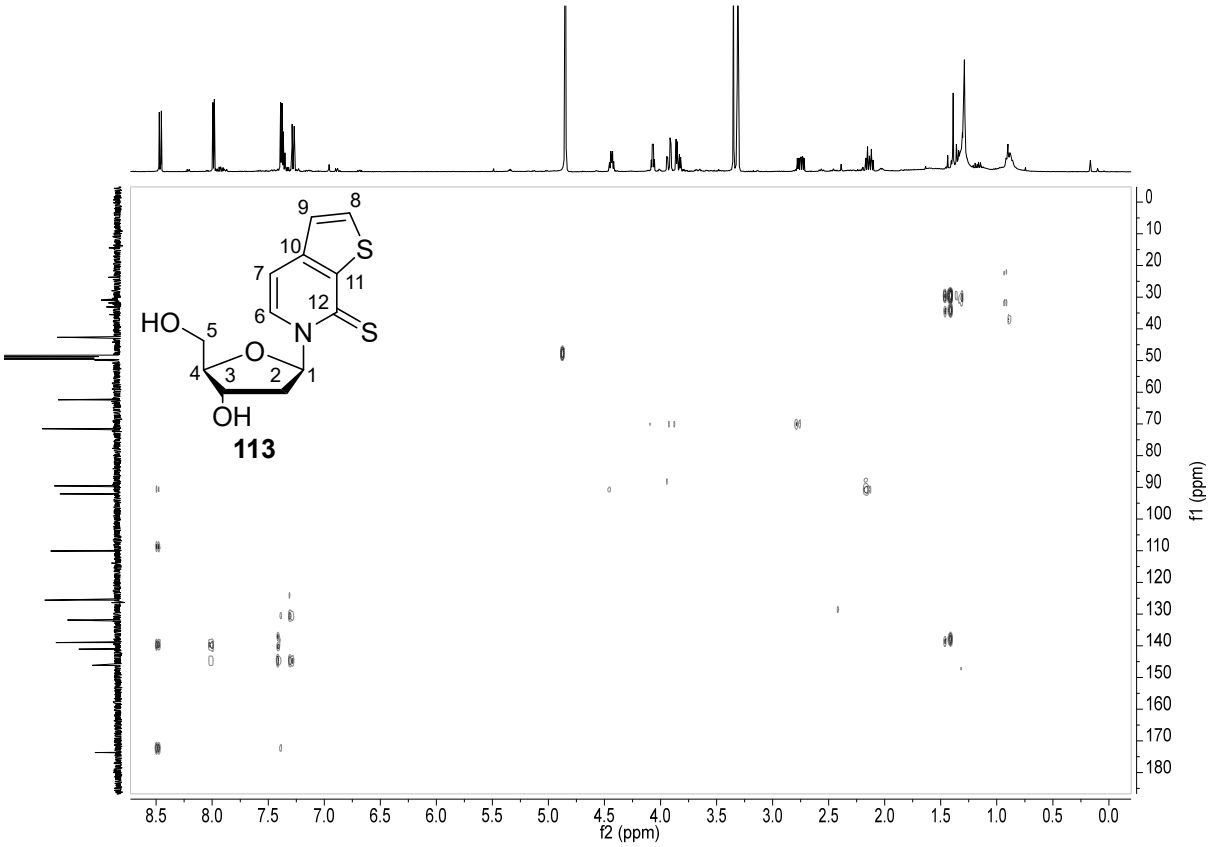


Figure 200. HMBC-NMR spectrum of compound **113**.

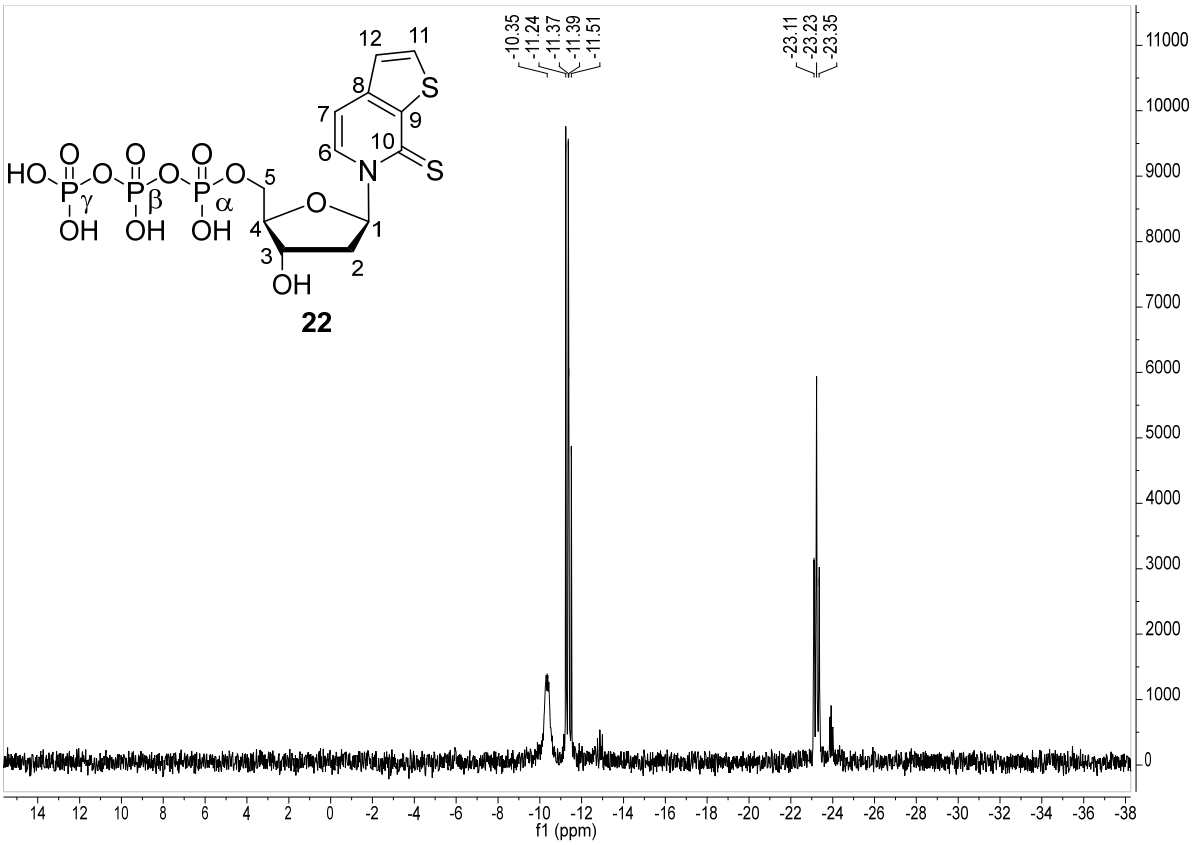


Figure 201. ^{31}P-NMR spectrum (D_2O, 162 MHz) of triphosphate **22**.

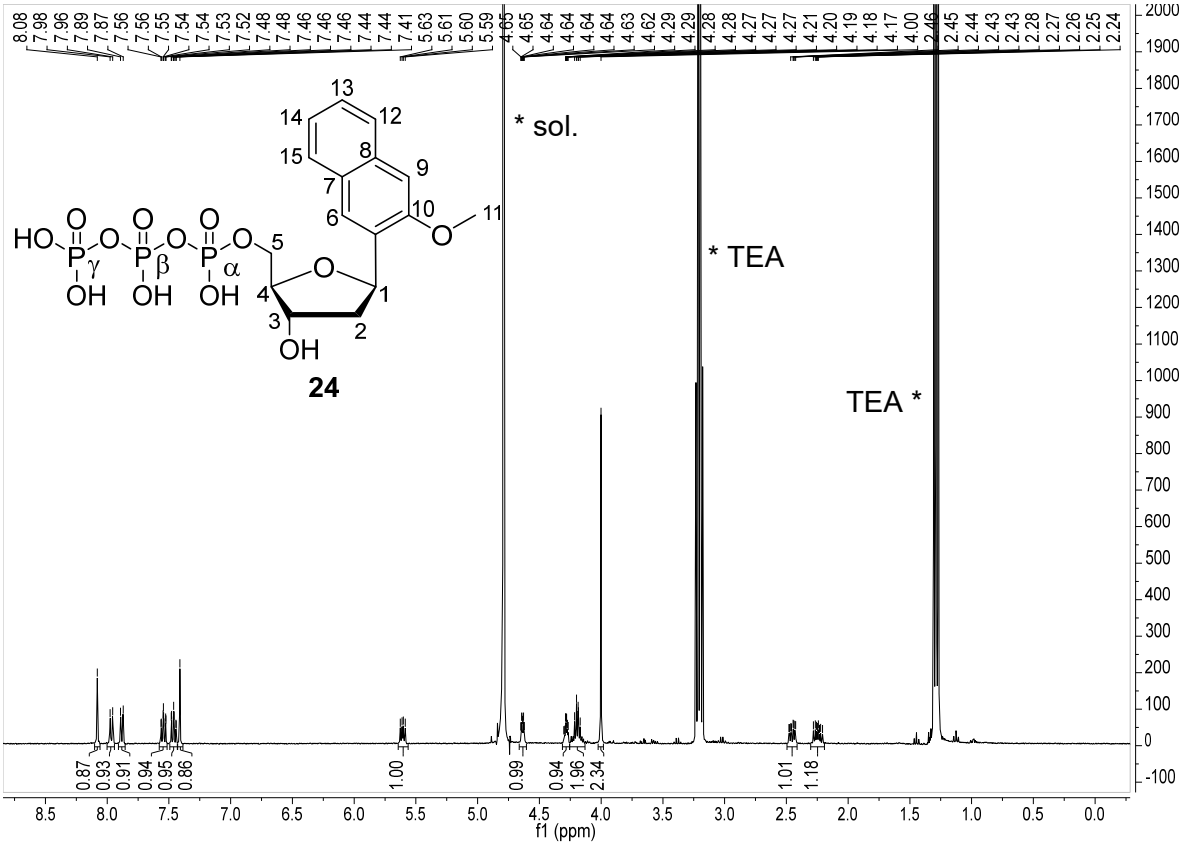


Figure 202. ^{1}H-NMR spectrum (D_2O, 400 MHz) of triphosphate **24**.

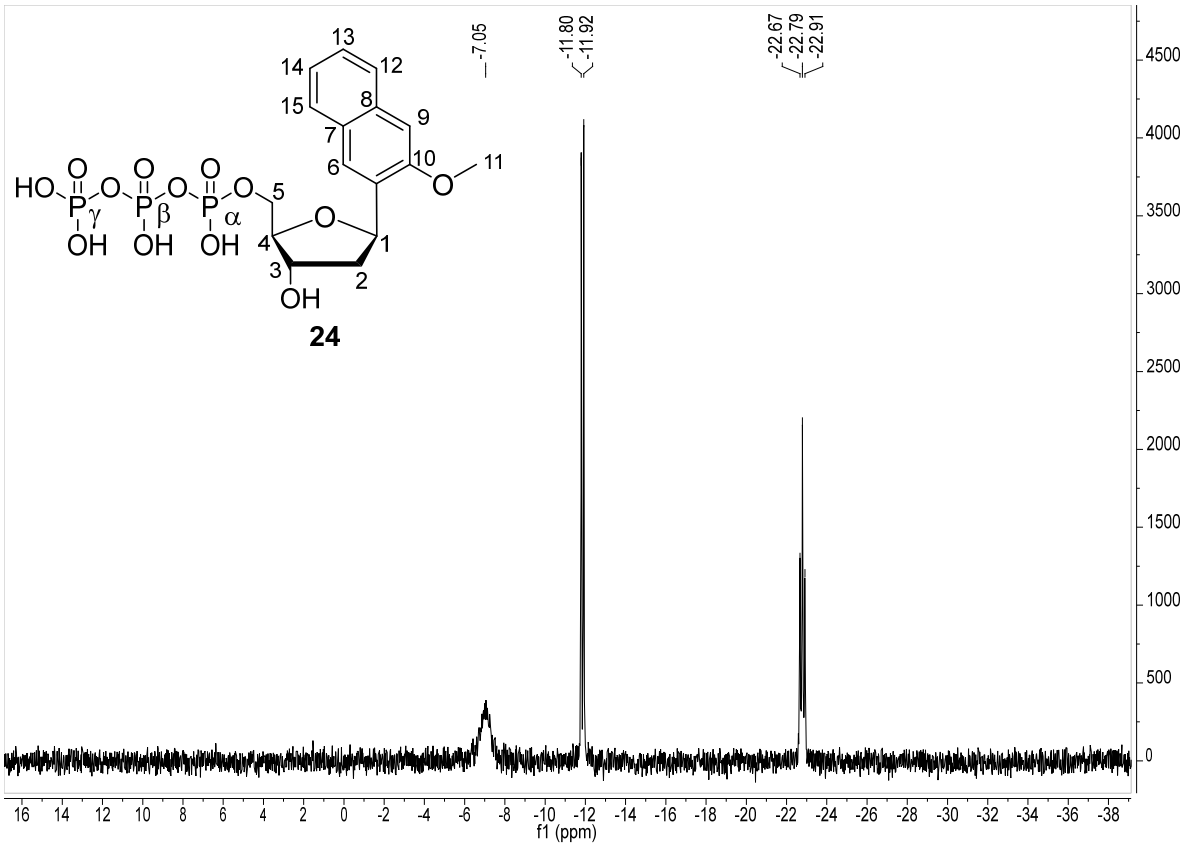


Figure 203. ^{31}P-NMR spectrum (D_2O, 162 MHz) of triphosphate **24**.

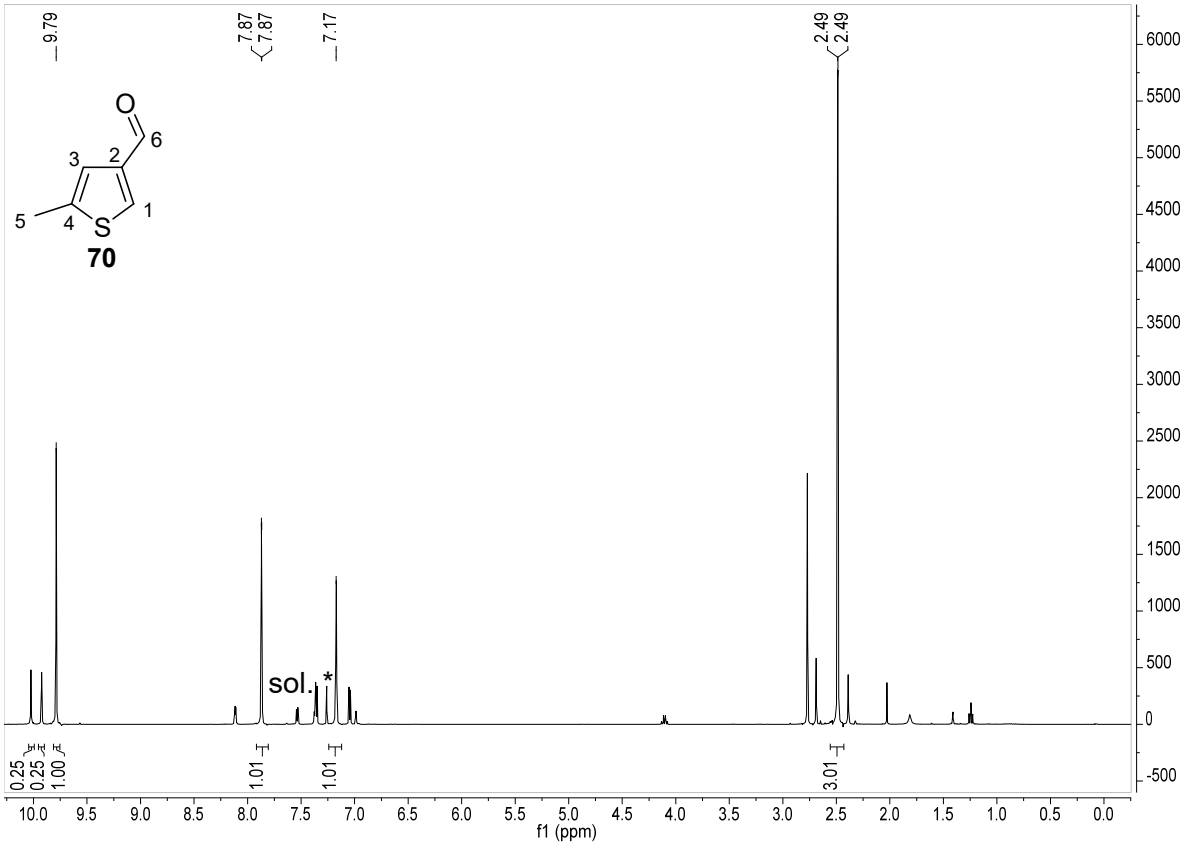


Figure 204. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of aldehyde **70**.

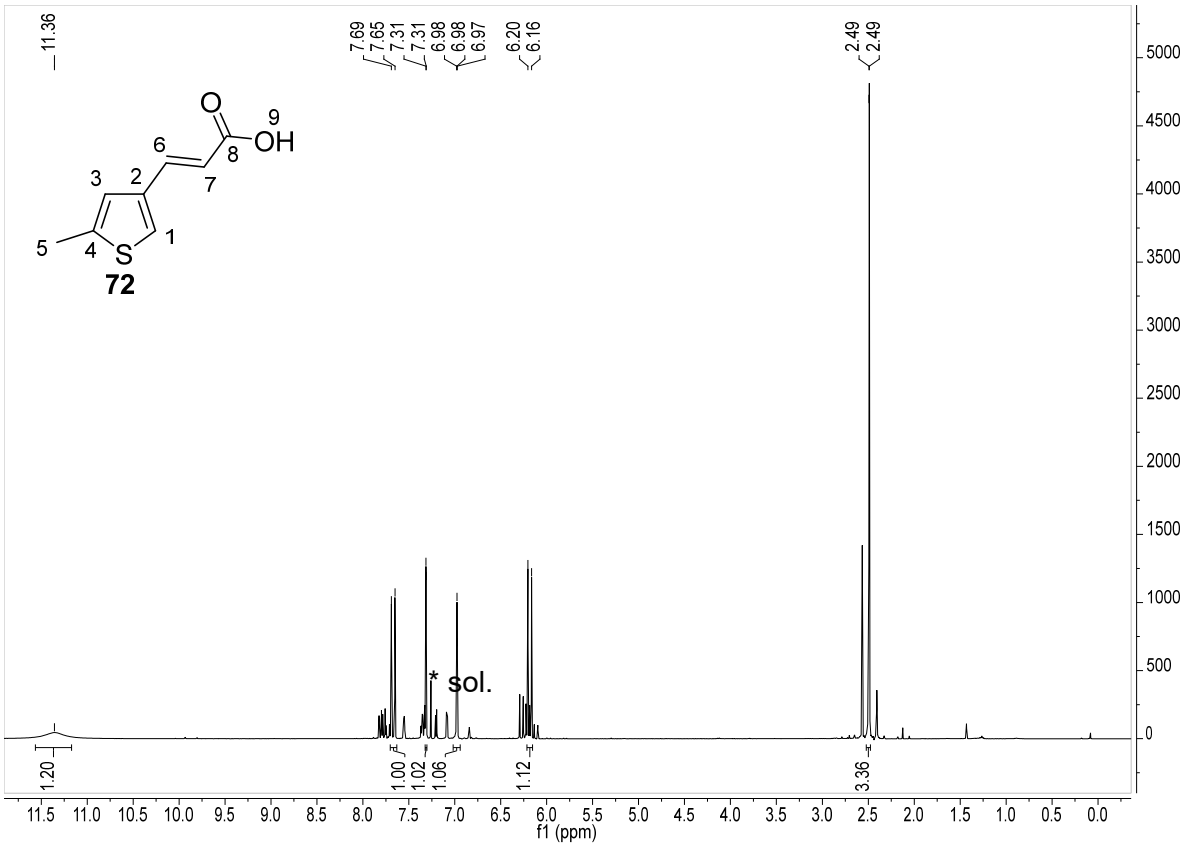


Figure 205. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of acrylic acid **72**.

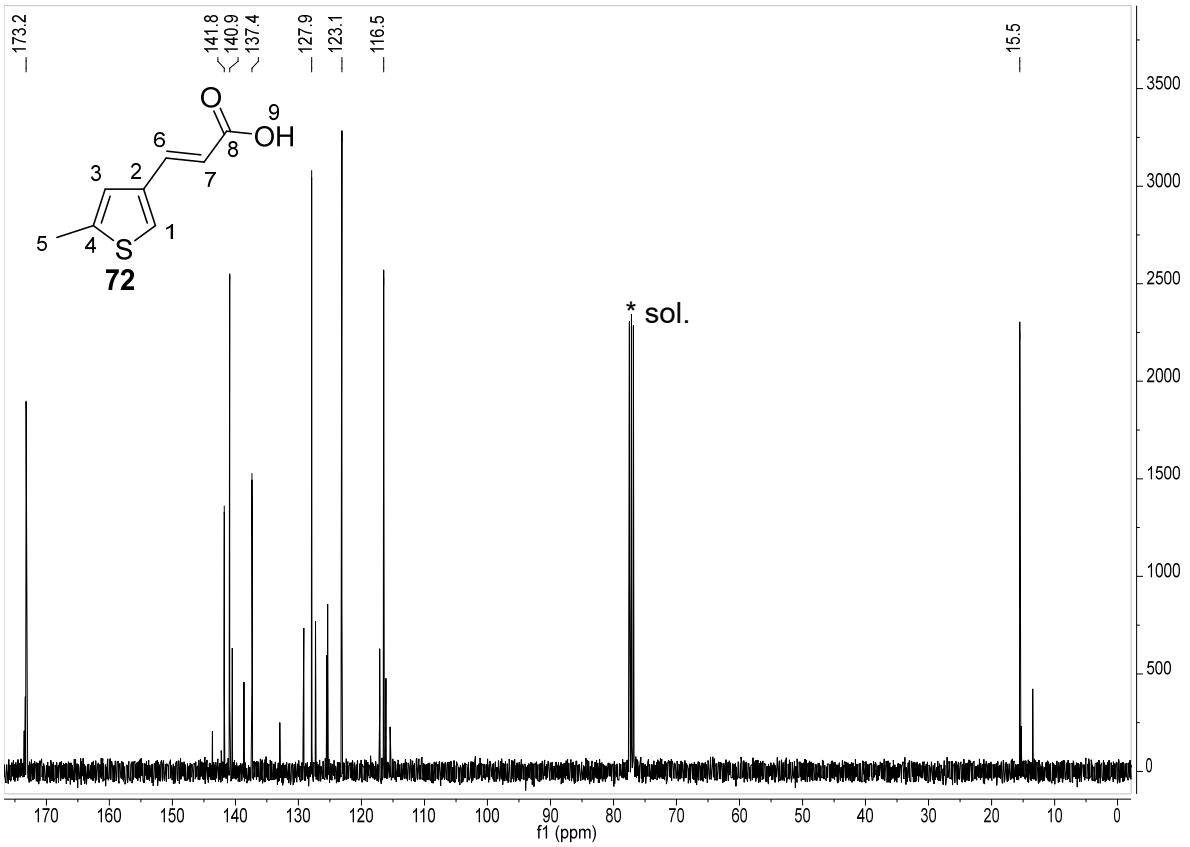


Figure 206. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of acrylic acid **72**.

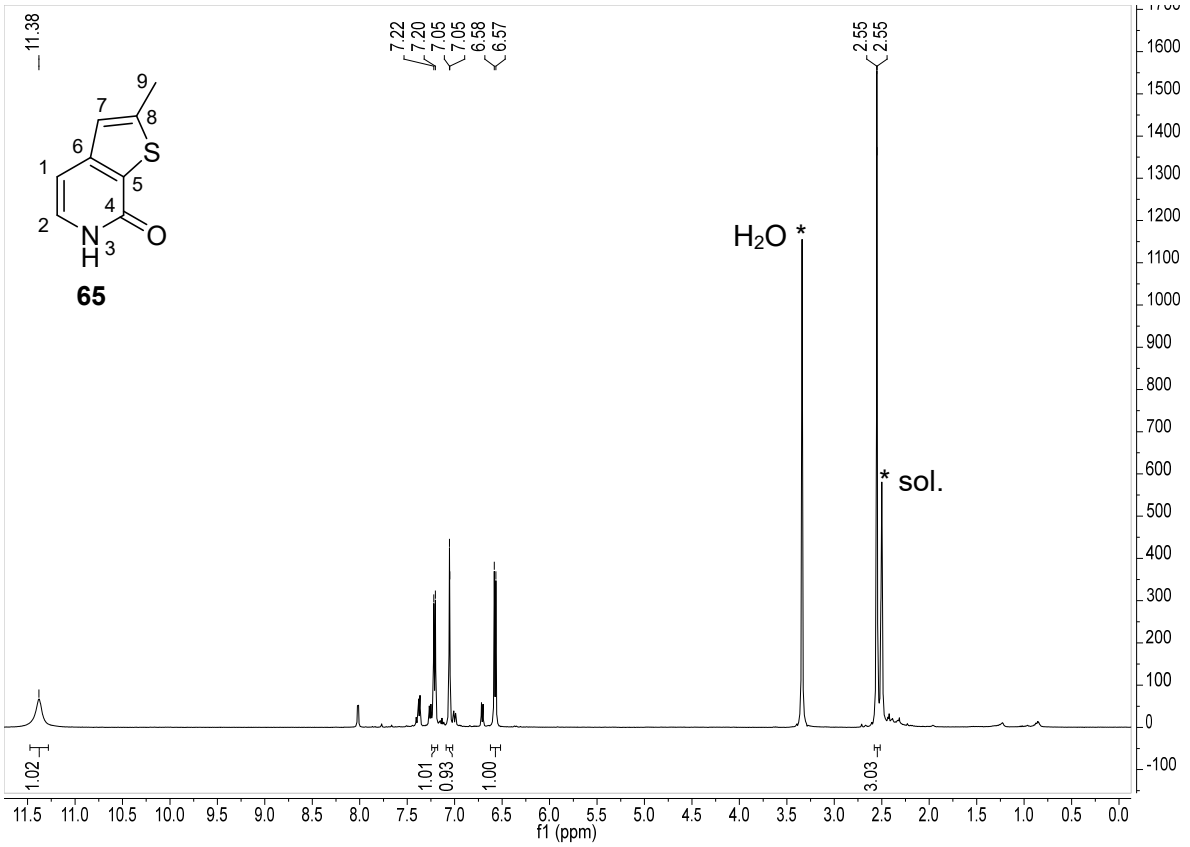


Figure 207. ^{1}H-NMR spectrum (DMSO-d_6, 400 MHz) of pyridone derivative **65**.

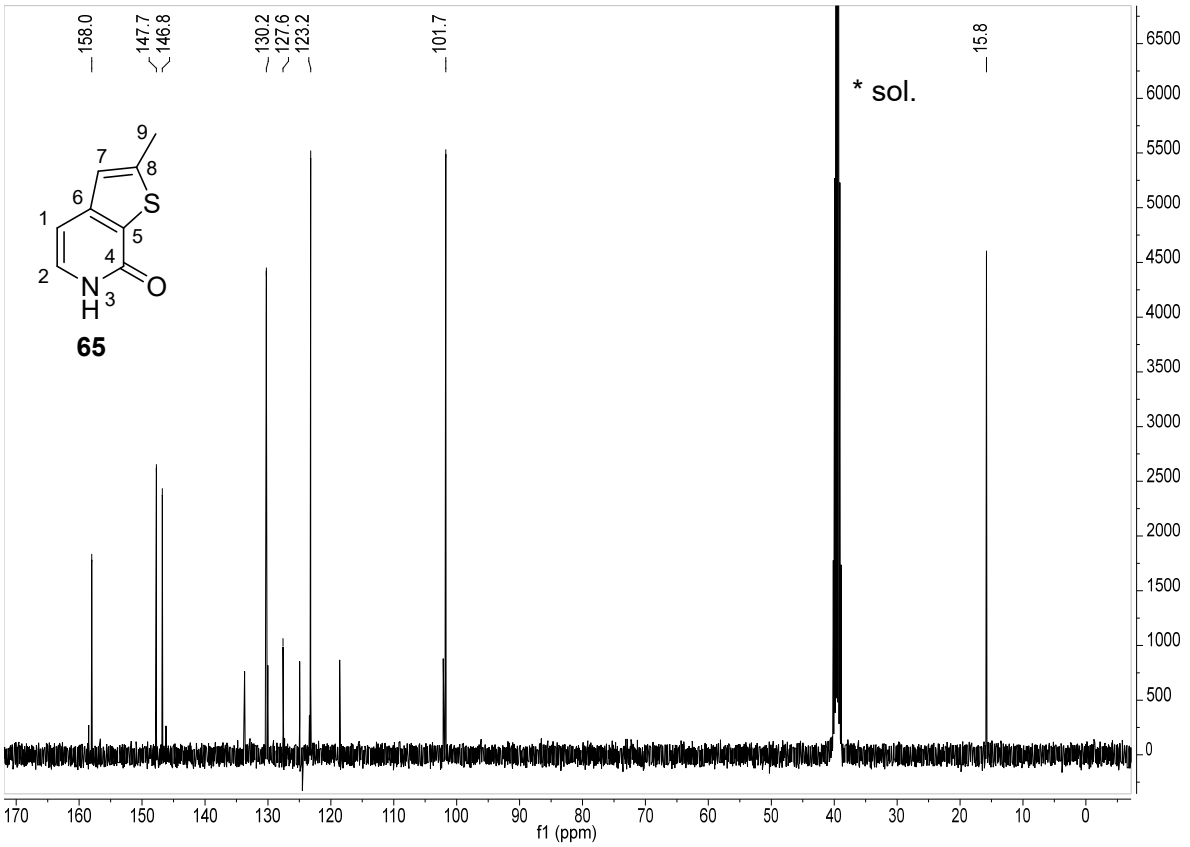


Figure 208. ^{13}C-NMR spectrum (DMSO-d_6, 101 MHz) of pyridone derivative **65**.

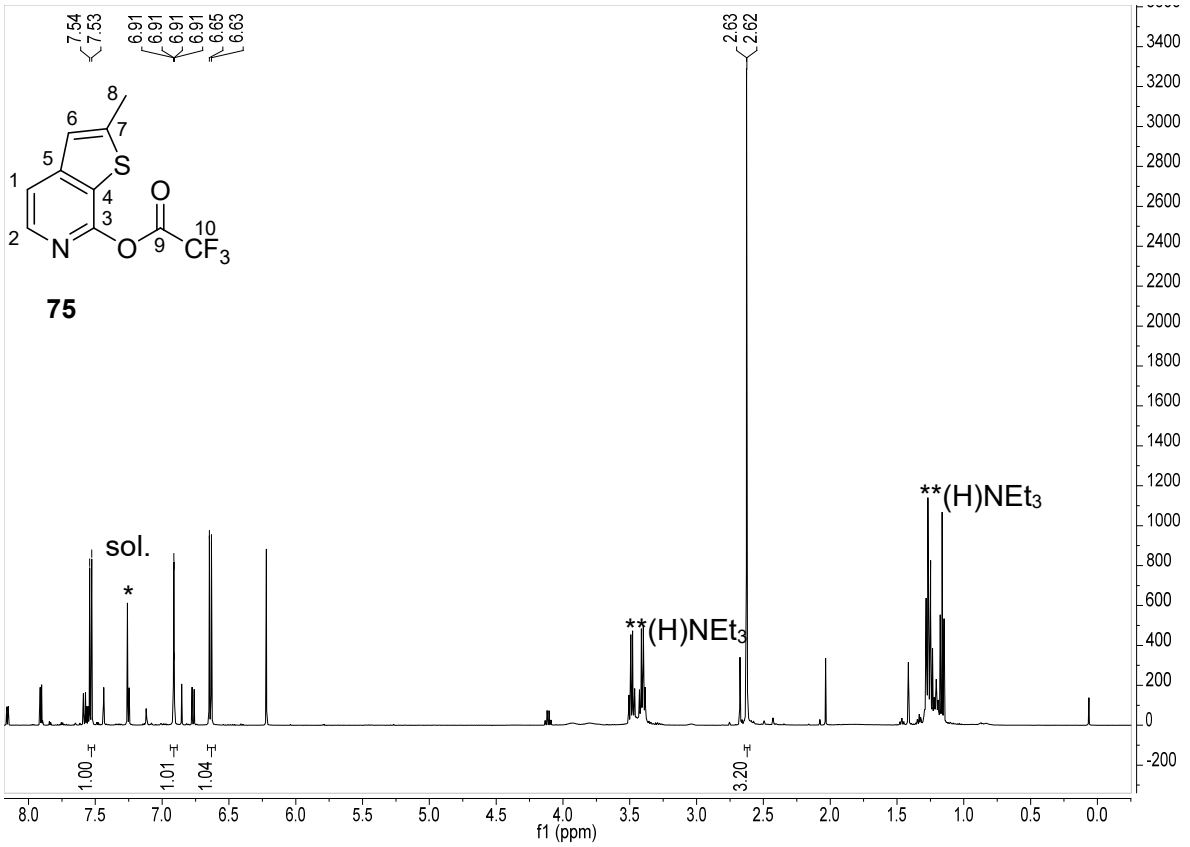


Figure 209. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **75**.

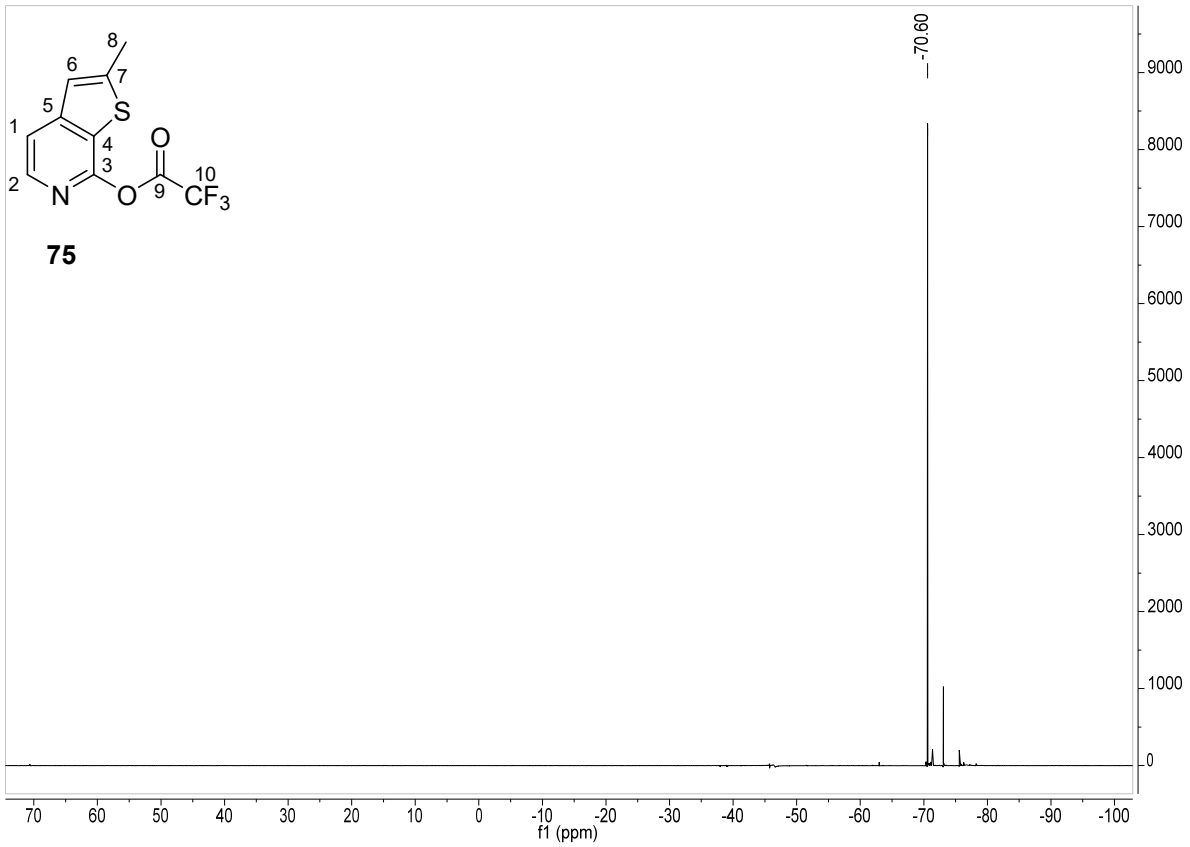
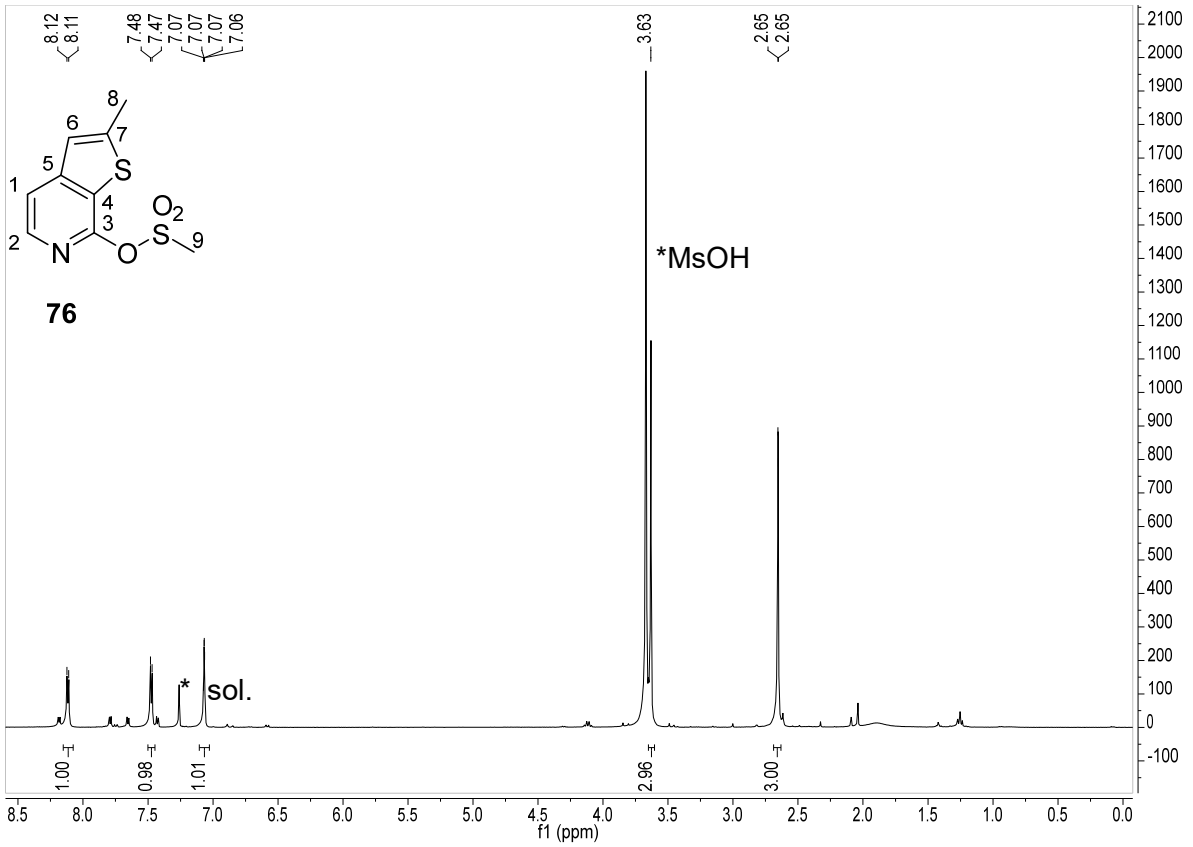
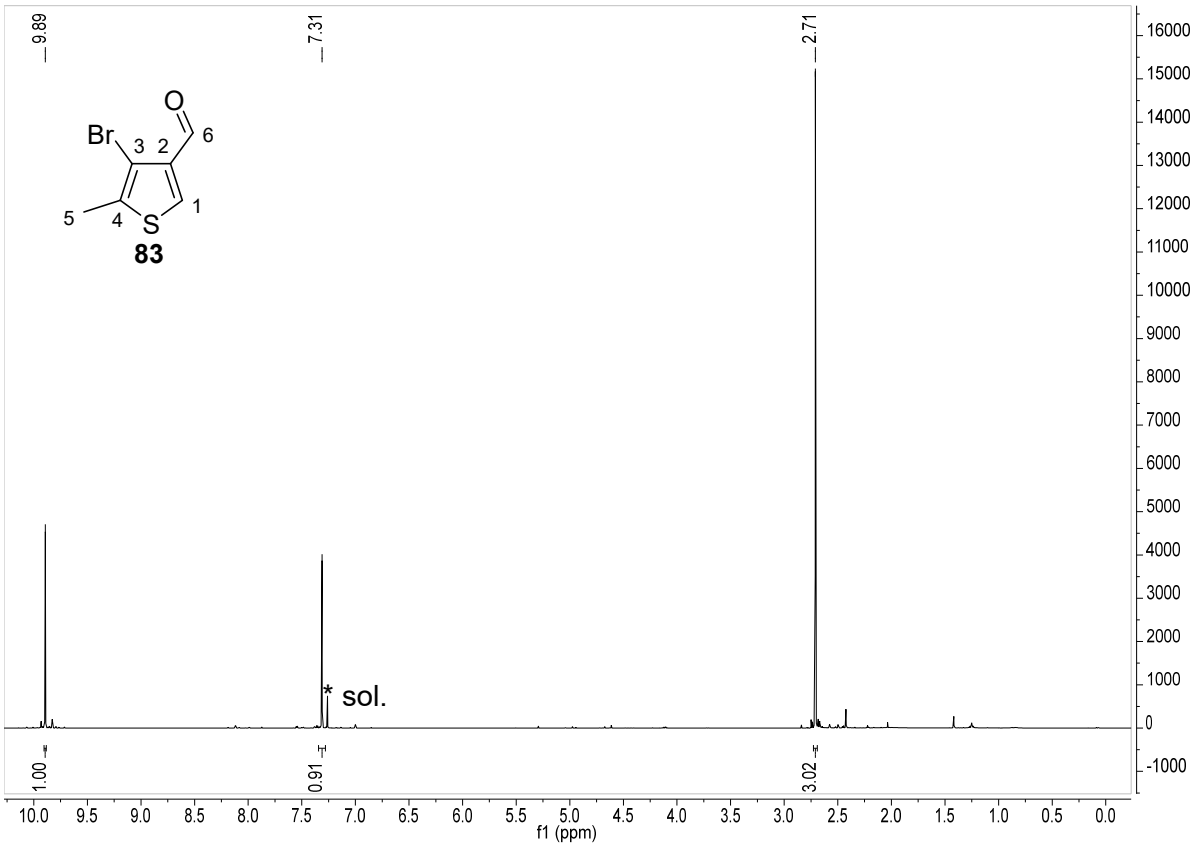


Figure 210. ^{19}F-NMR ($CDCl_3$, 470 MHz) of compound **75**.



76

Figure 211. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **76**.



83

Figure 212. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **83**.

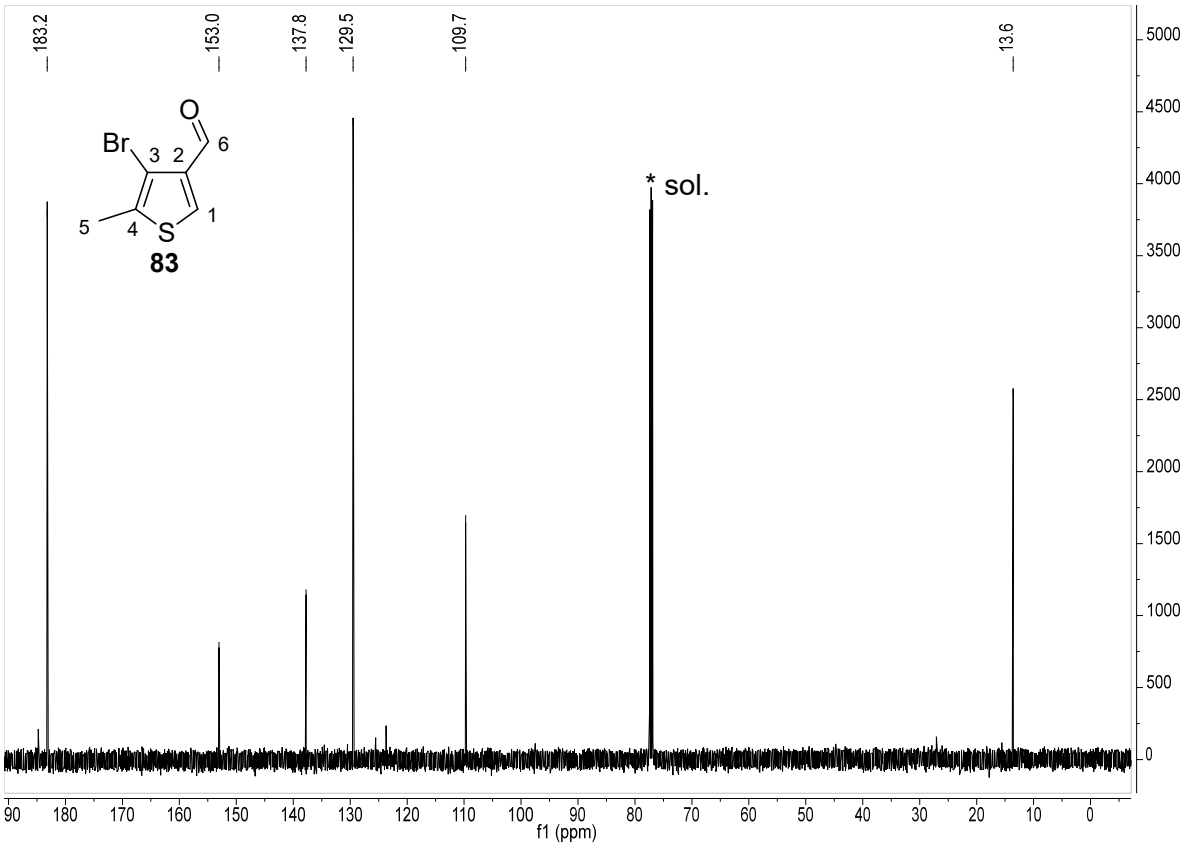


Figure 213. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of compound **83**.

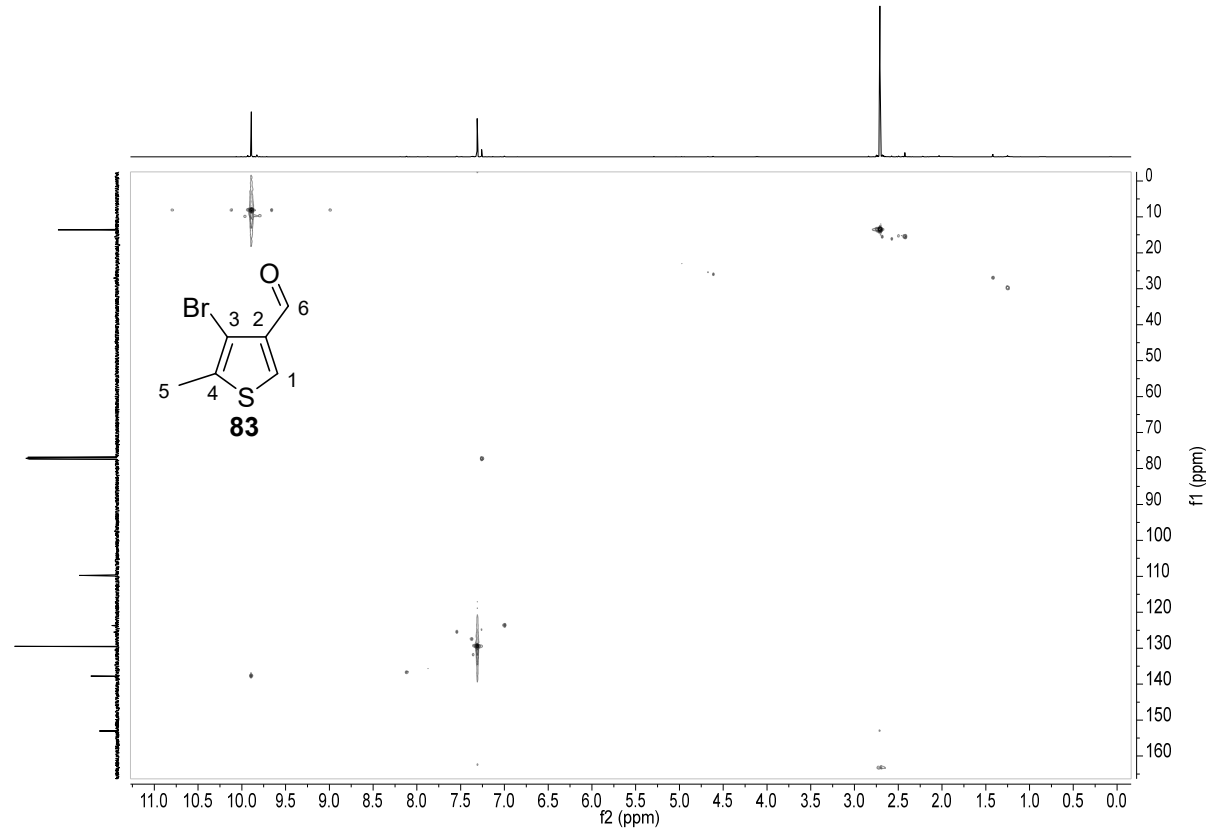


Figure 214. HSQC-NMR spectrum of compound **83**.

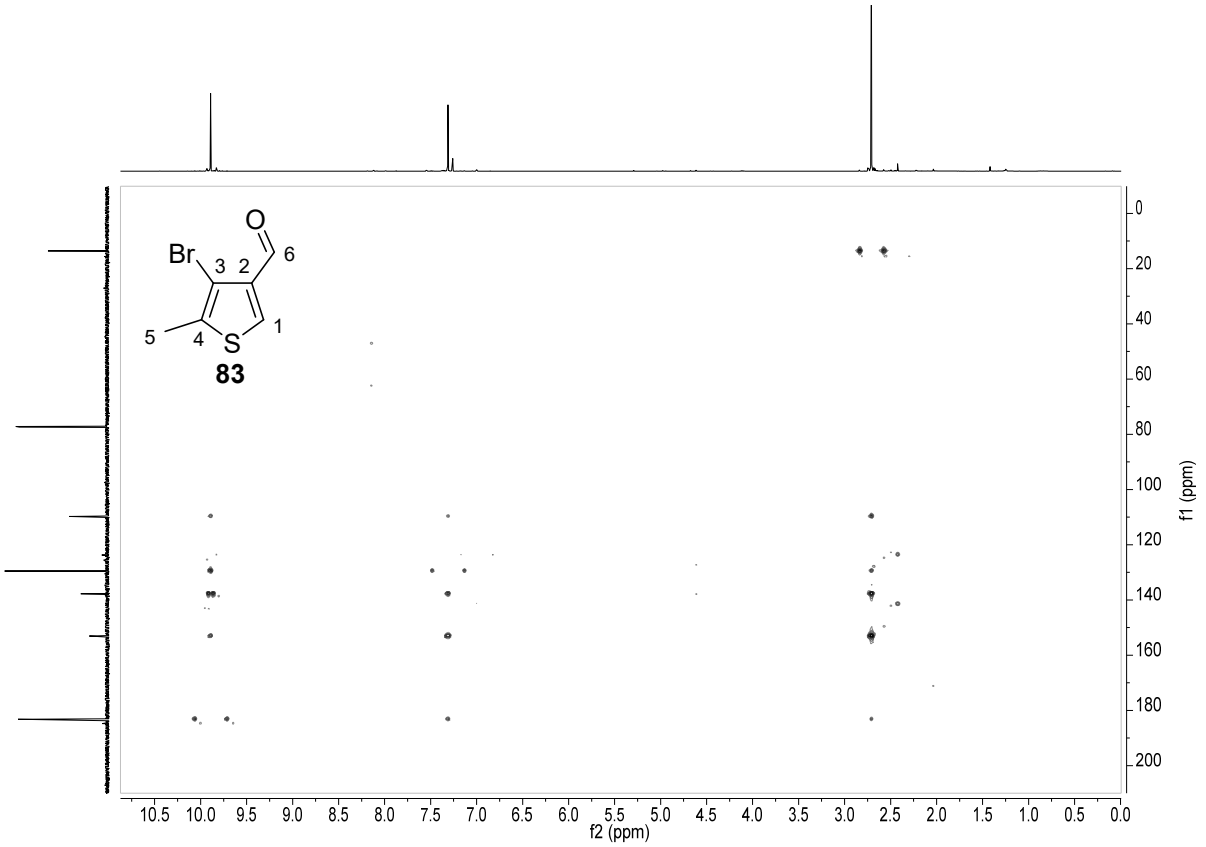


Figure 215. HMBC-NMR spectrum of compound **83**.

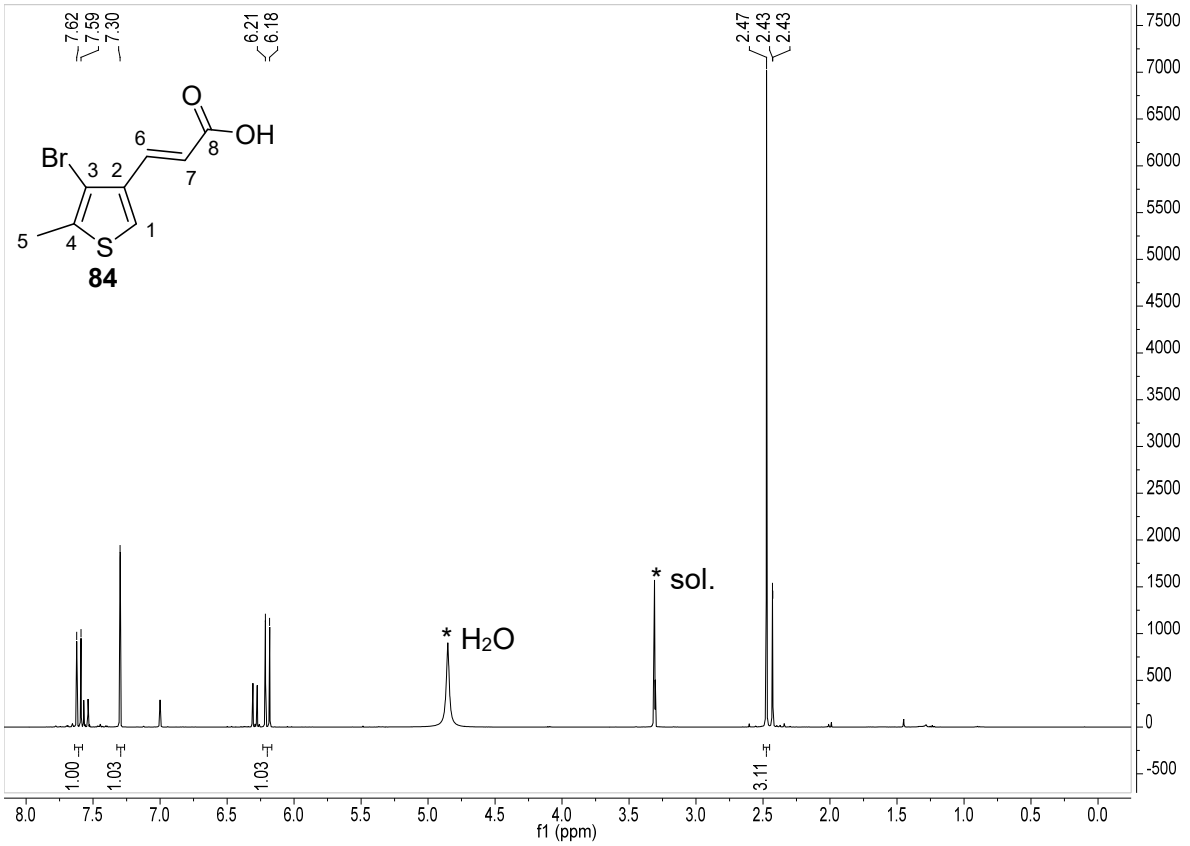


Figure 216. ^{1}H-NMR spectrum (CD_3OD, 500 MHz) of acrylic acid **84**.

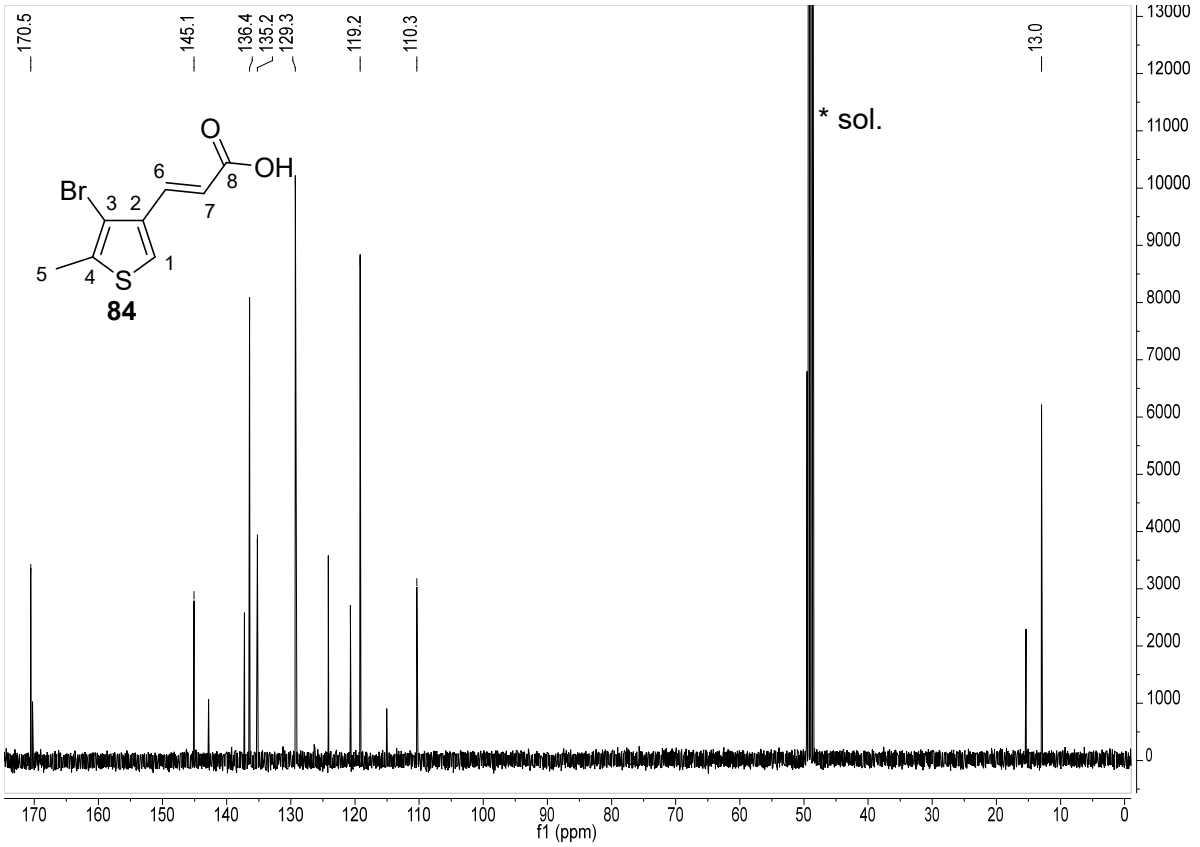


Figure 217. ^{13}C-NMR spectrum (CD_3OD, 126 MHz) of acrylic acid **84**.

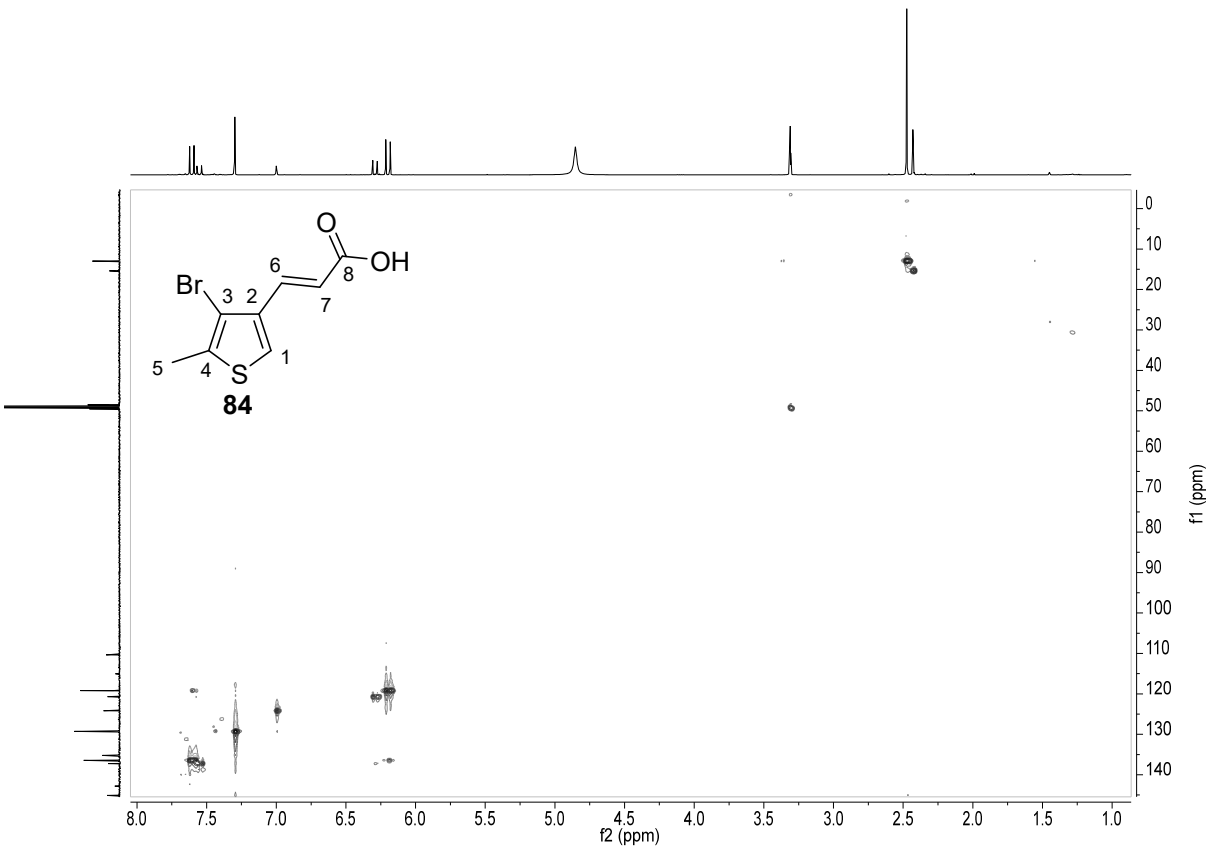


Figure 218. HSQC-NMR spectrum of acrylic acid **84**.

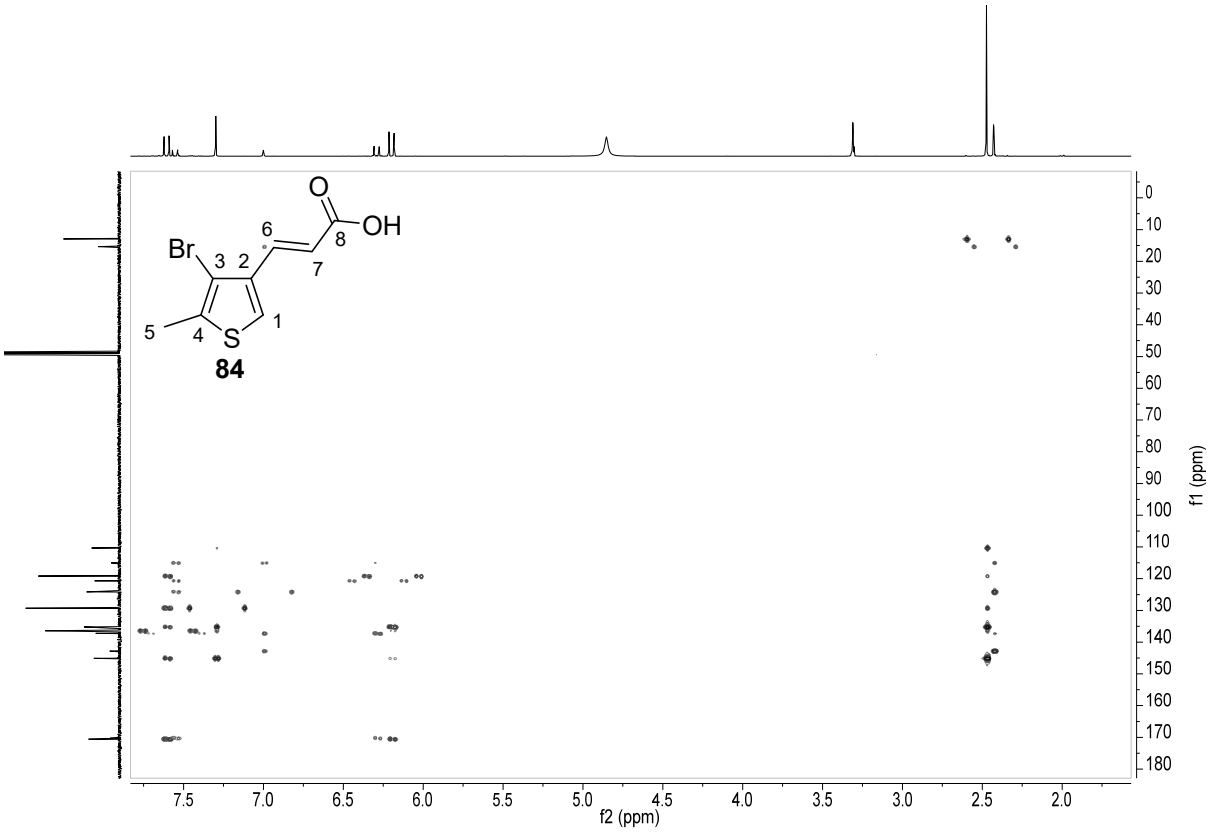


Figure 219. HMBC-NMR spectrum of acrylic acid **84**.

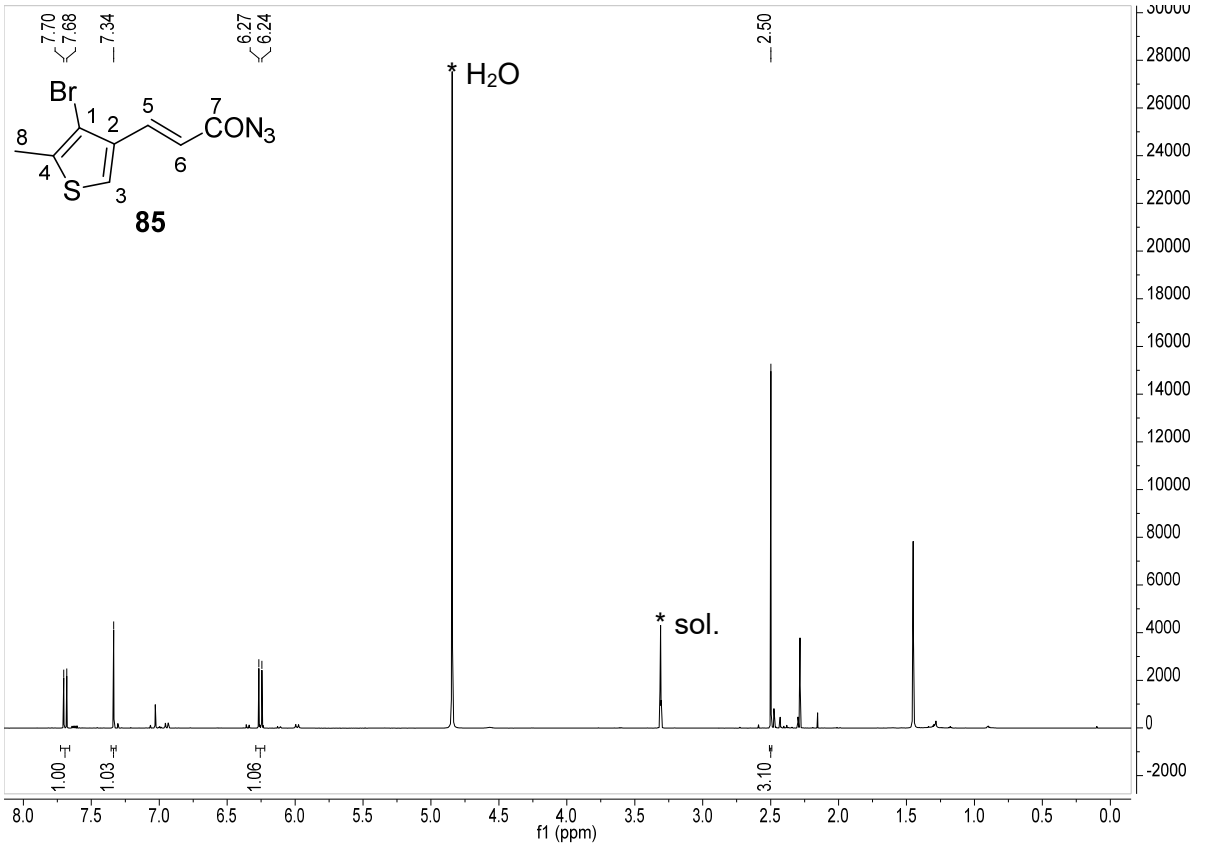


Figure 220. ^{1}H-NMR spectrum (CD_3OD, 700 MHz) of acid aczide **85**.

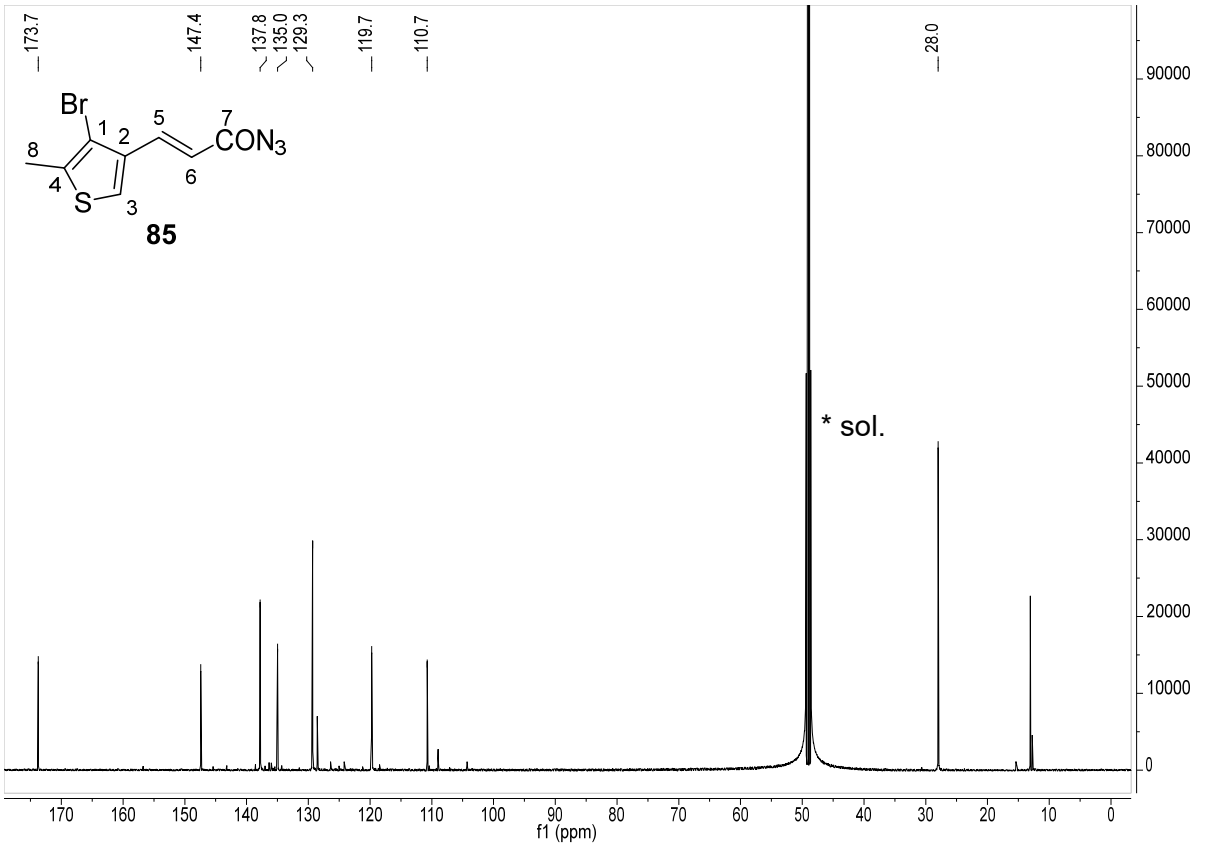


Figure 221. ^{13}C-NMR spectrum (CD_3OD, 176 MHz) of acid azide **85**.

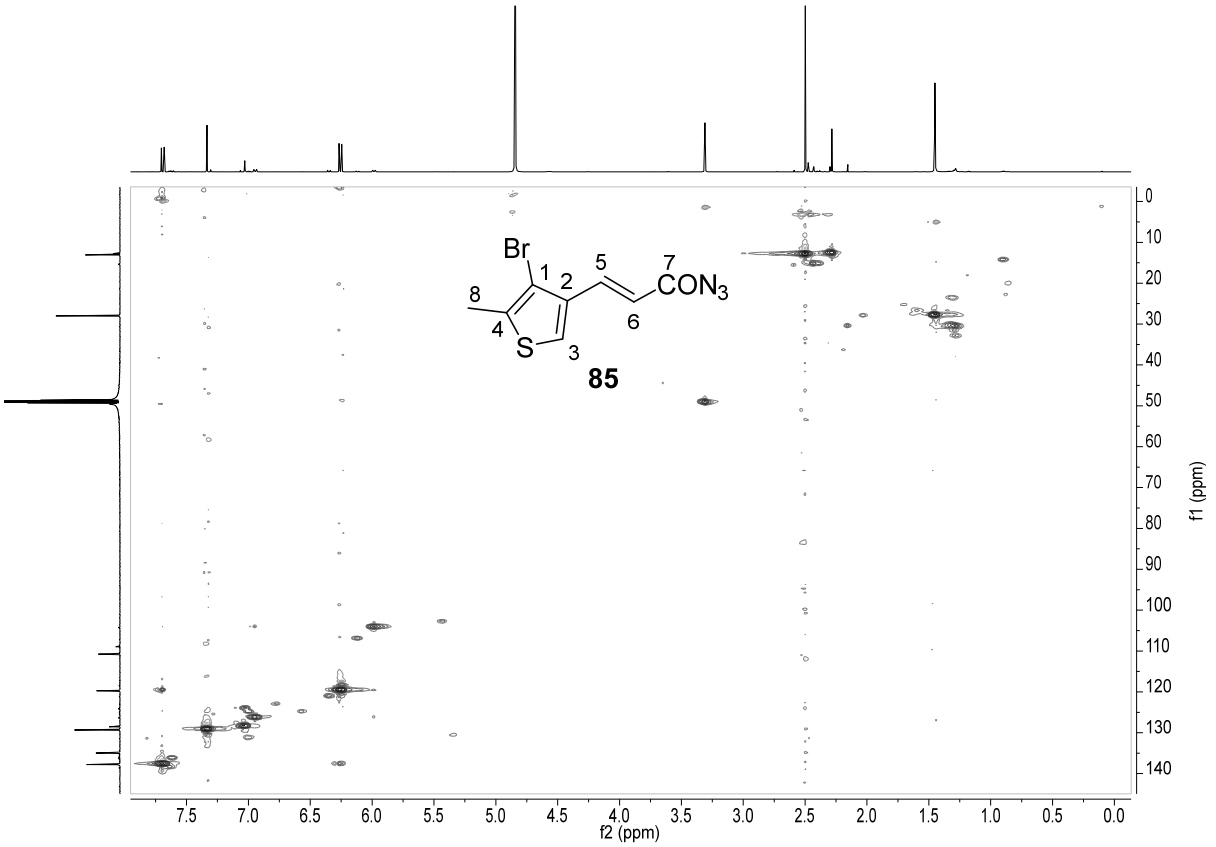


Figure 222. HSQC-NMR spectrum of acid azide **85**.

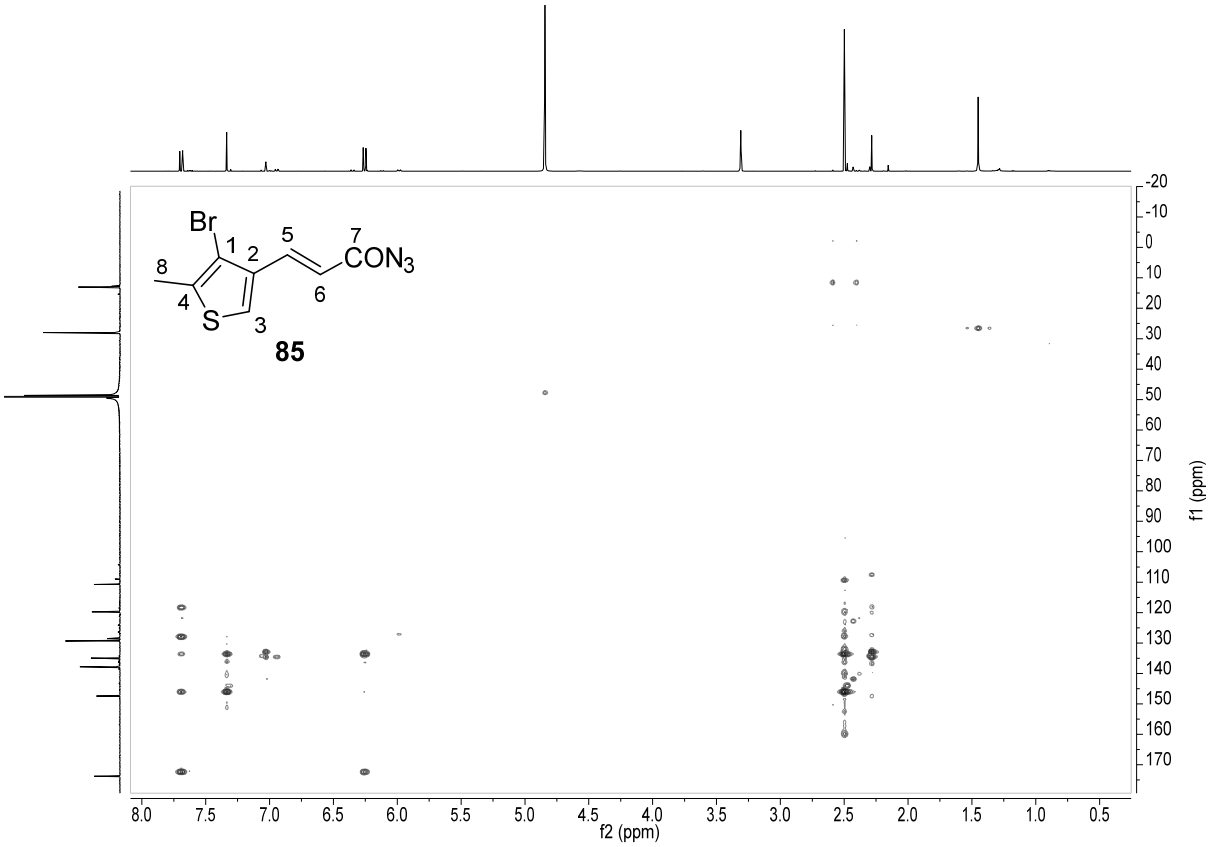


Figure 223. HMBC-NMR spectrum of acid azide **85**.

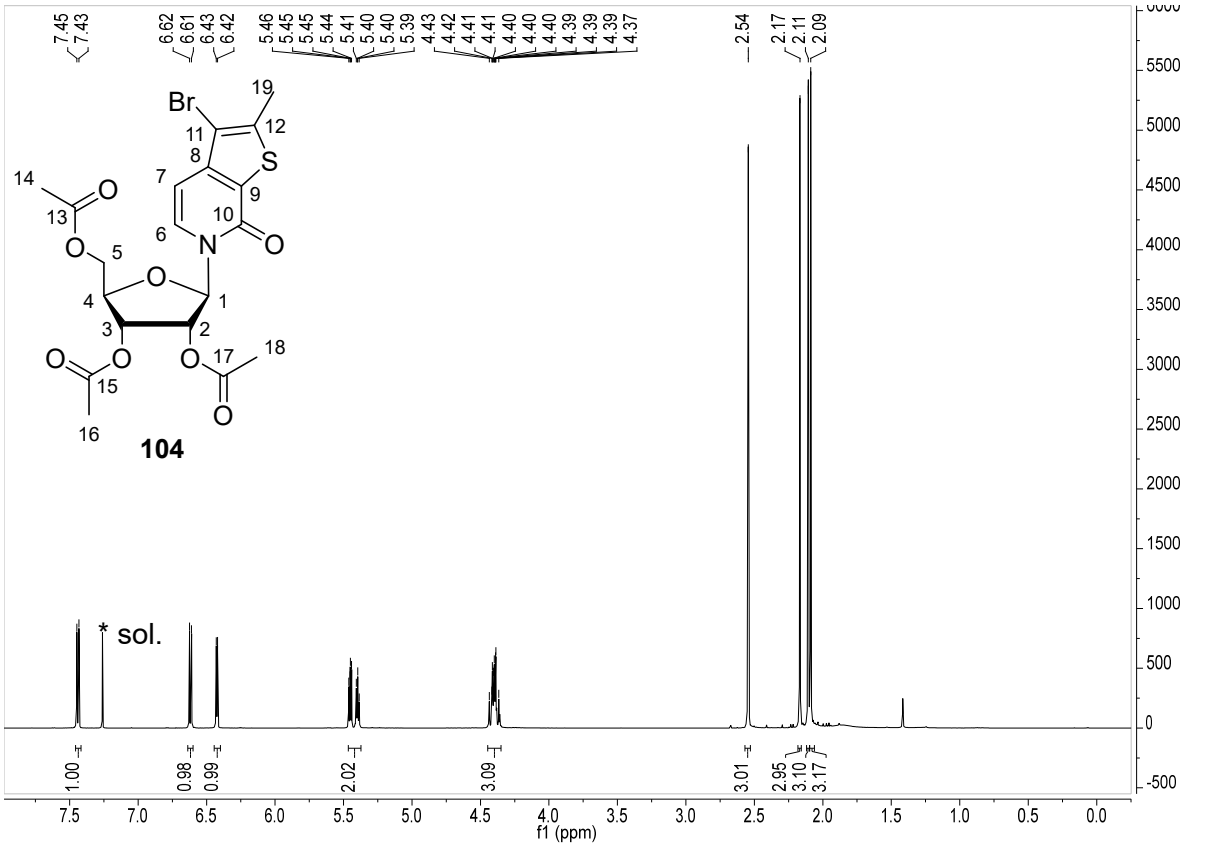


Figure 224. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **104**.

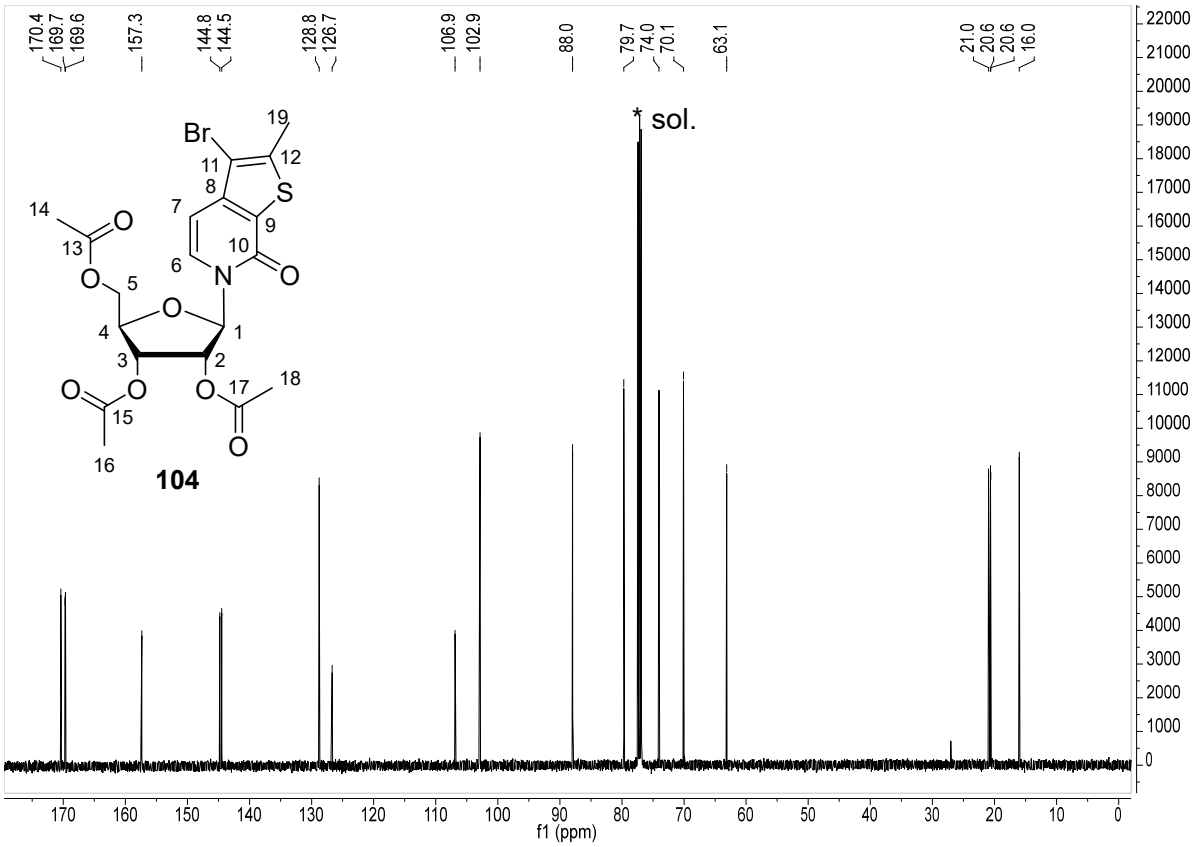


Figure 225. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of compound **104**.

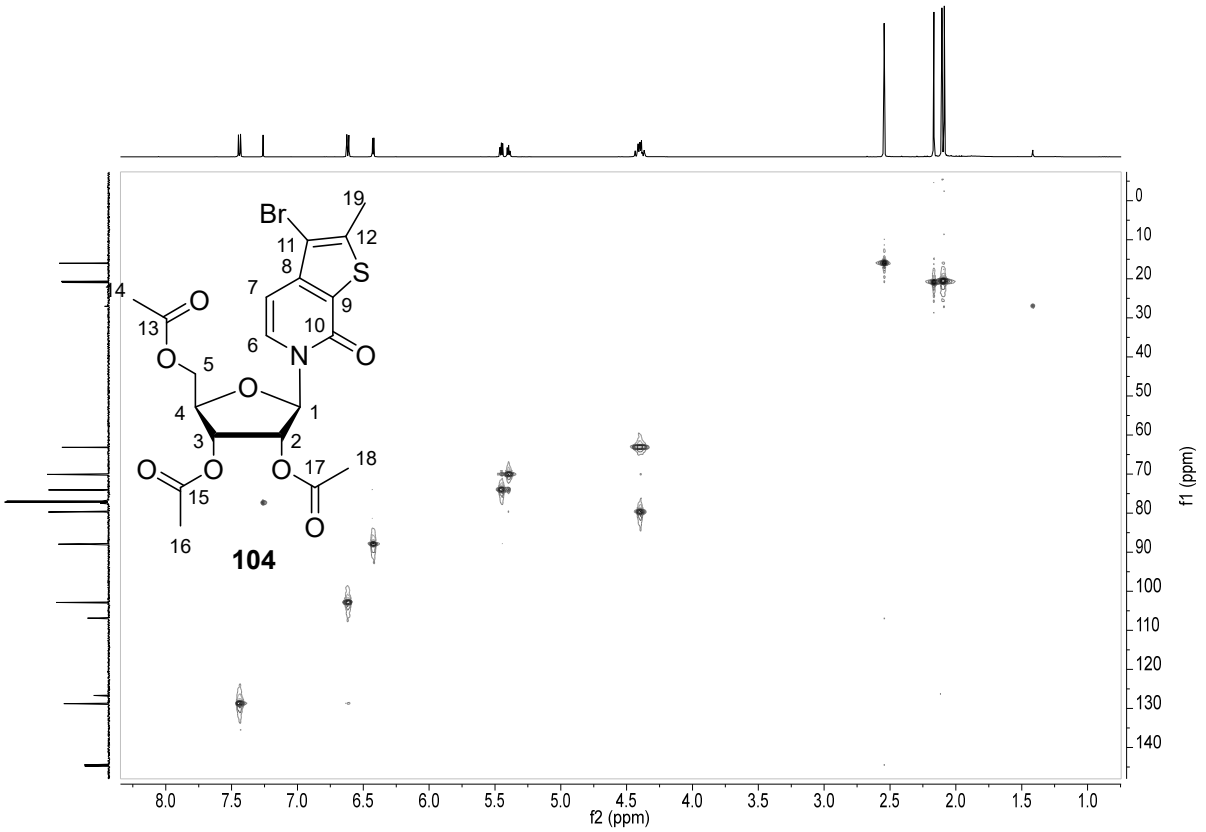


Figure 226. HSQC-NMR spectrum of compound **104**.

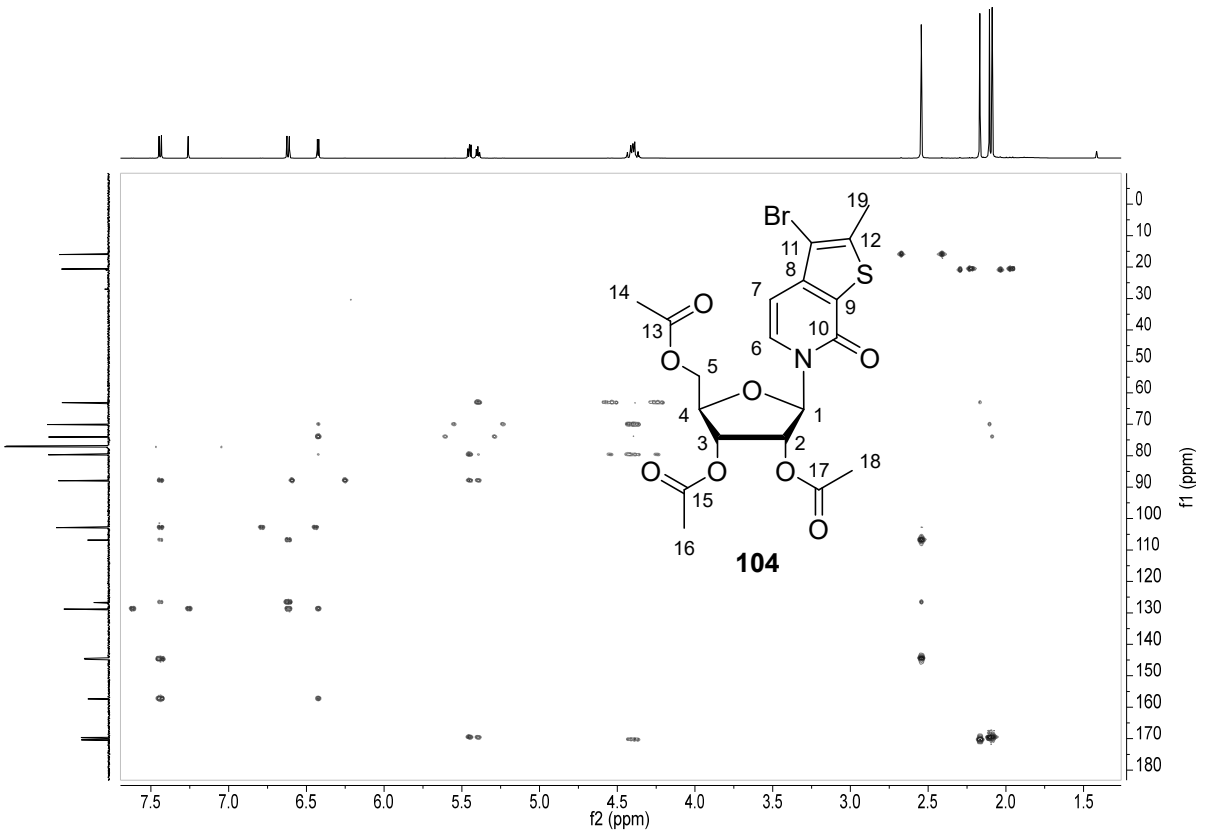


Figure 227. HMBC-NMR spectrum of compound **104**.

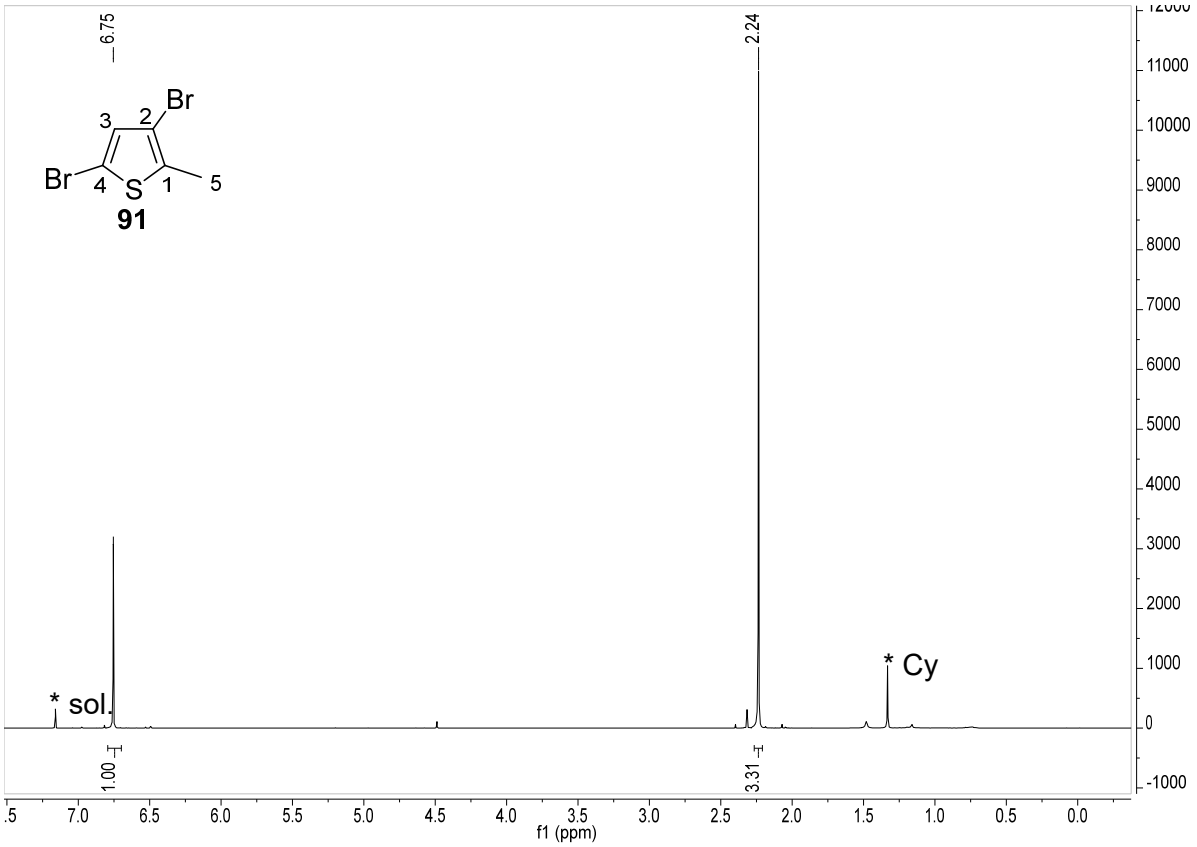


Figure 228. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of thiophene **91**.

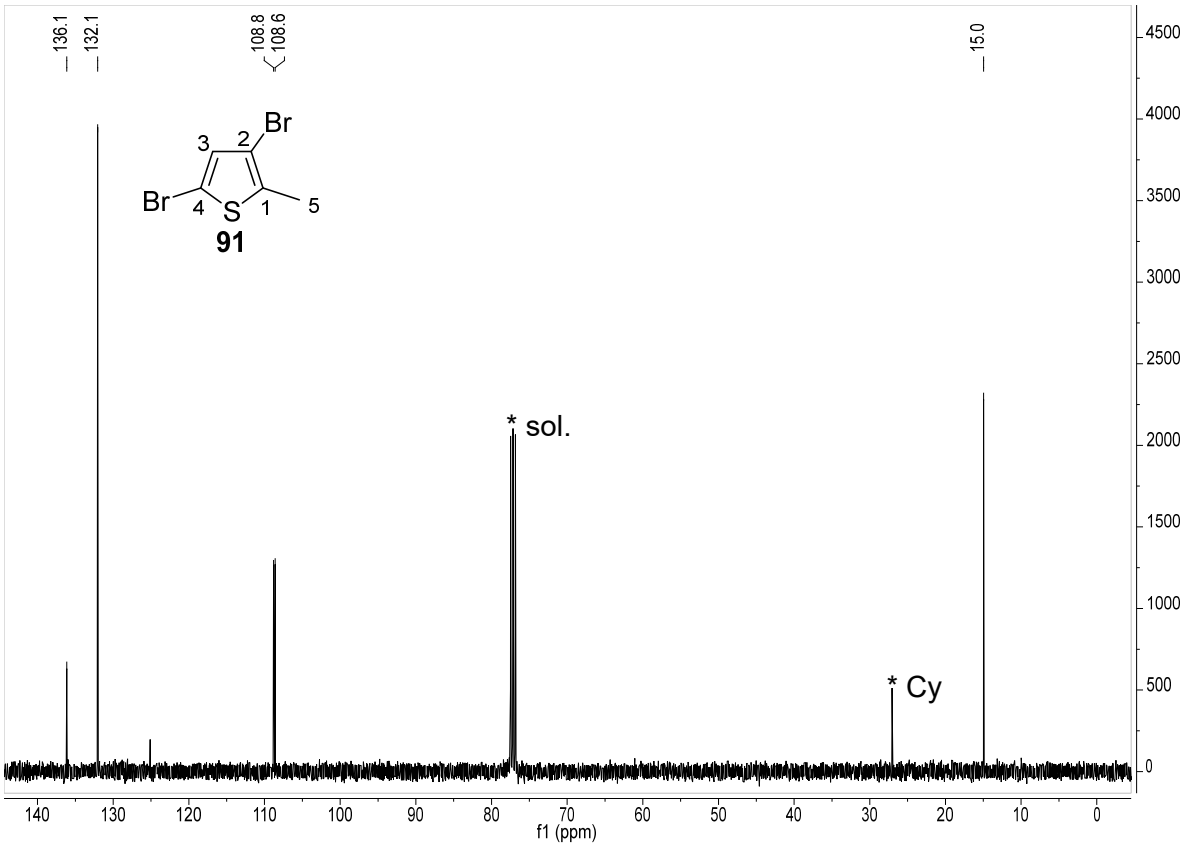


Figure 229. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of thiophene **91**.

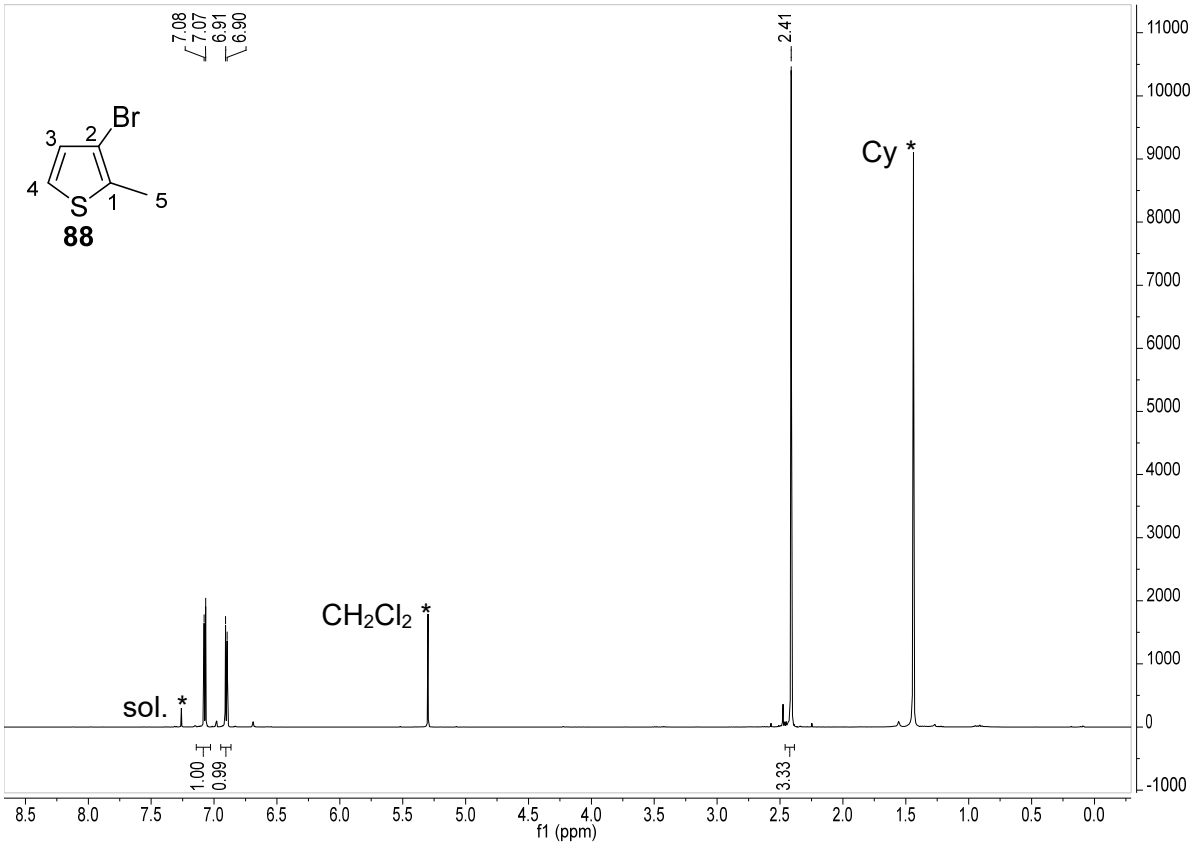


Figure 230. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of thiophene **88**.

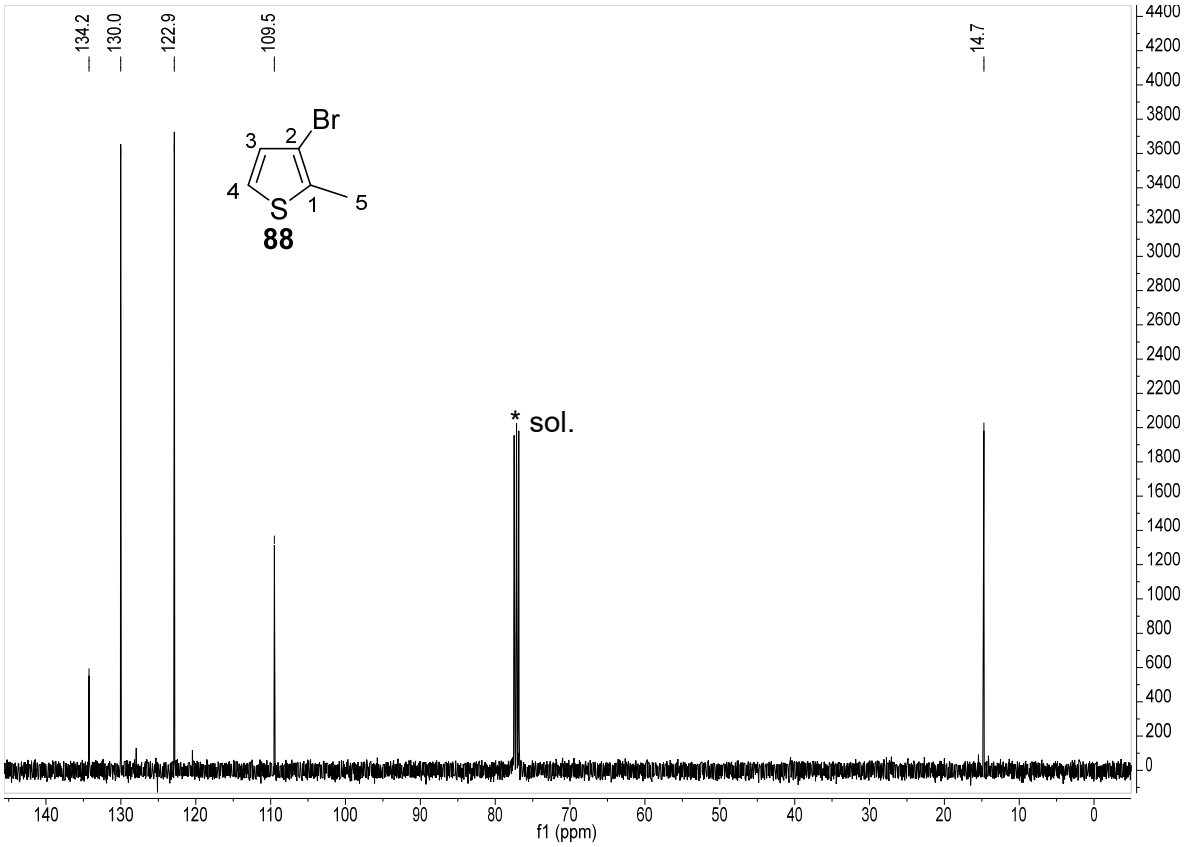


Figure 231. ^{13}C-NMR spectrum ($CDCl_3$, 400 MHz) of thiophene **88**.

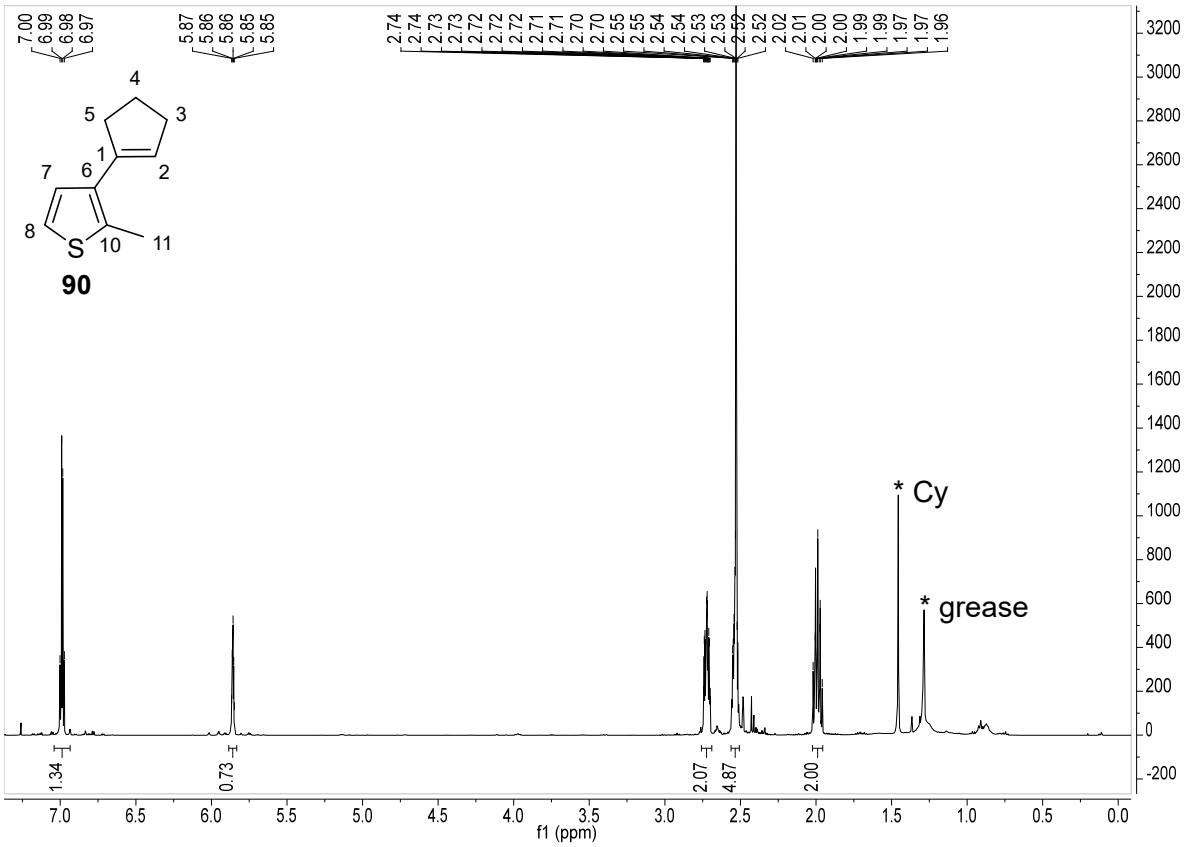


Figure 232. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of cyclopentene-thiophene adduct **90**.

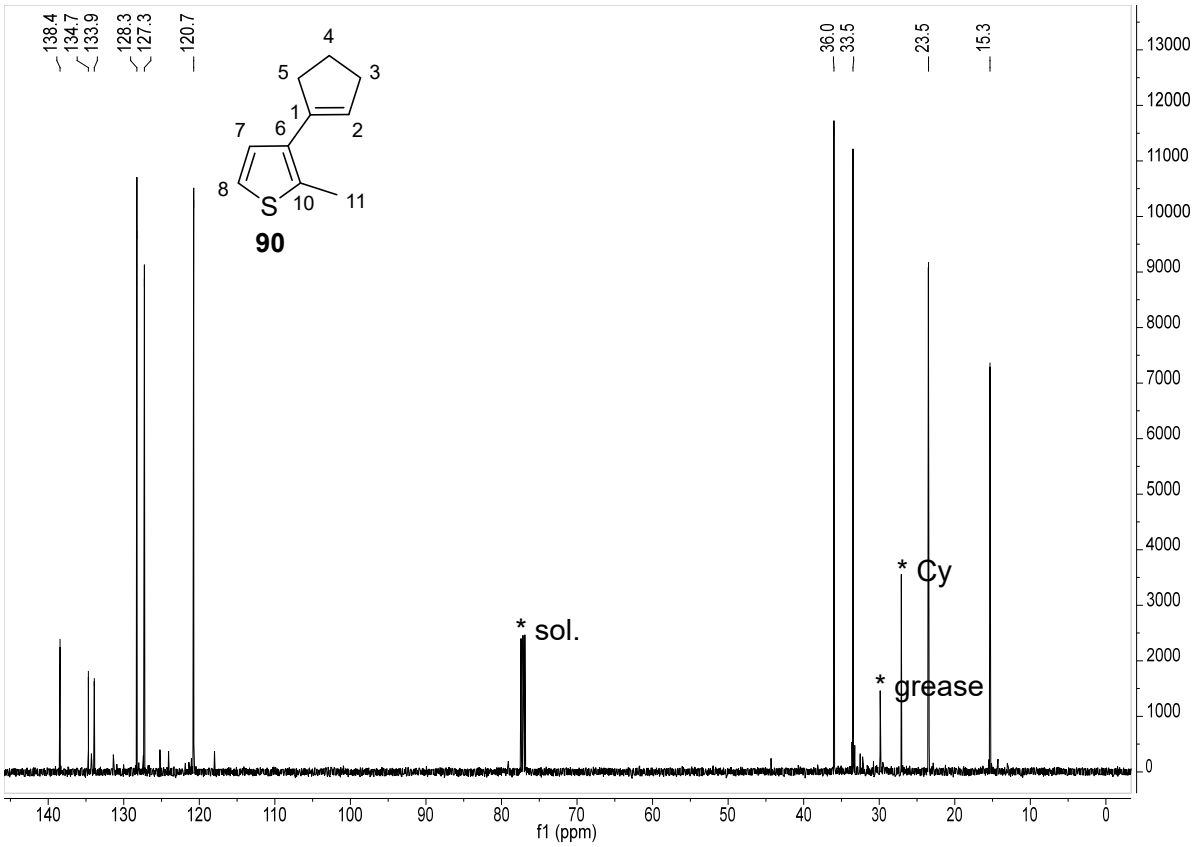


Figure 233. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of cyclopentene-thiophene adduct **90**.

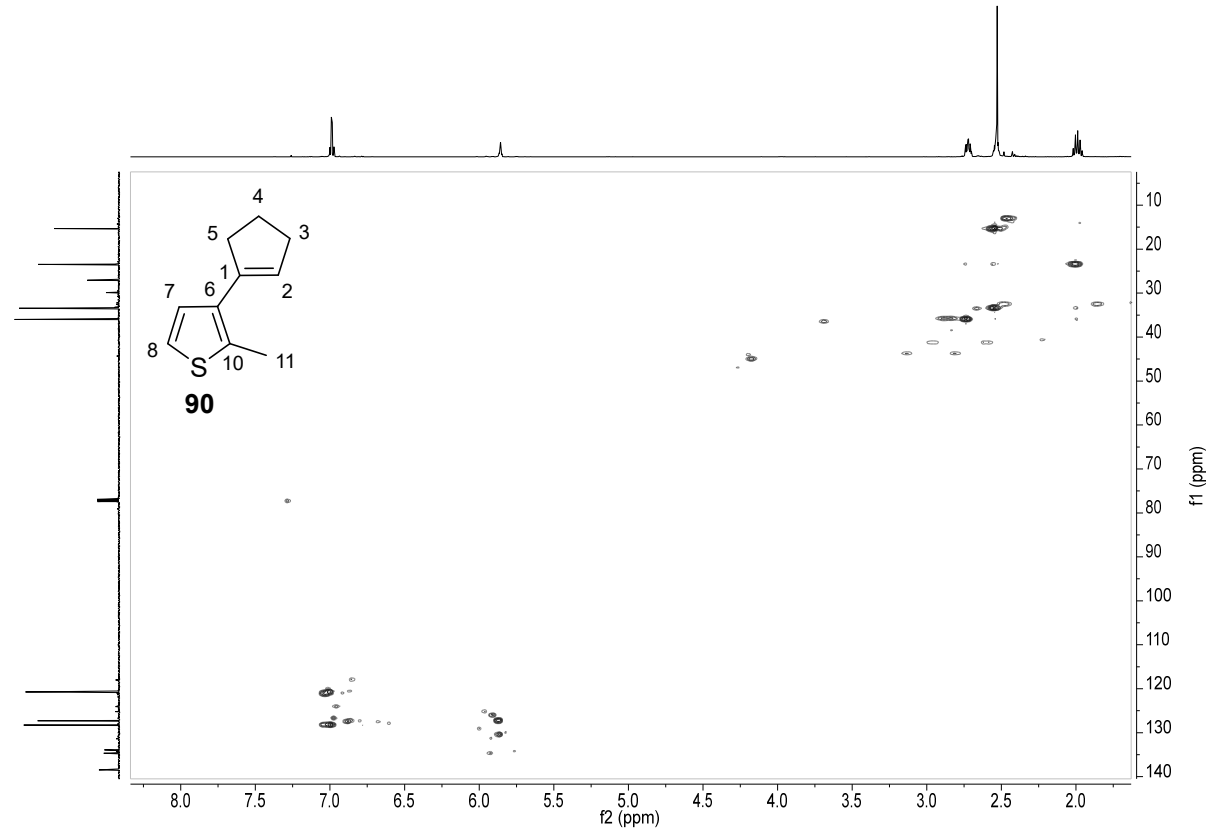


Figure 234. HSQC-NMR spectrum of cyclopentene-thiophene adduct **90**.

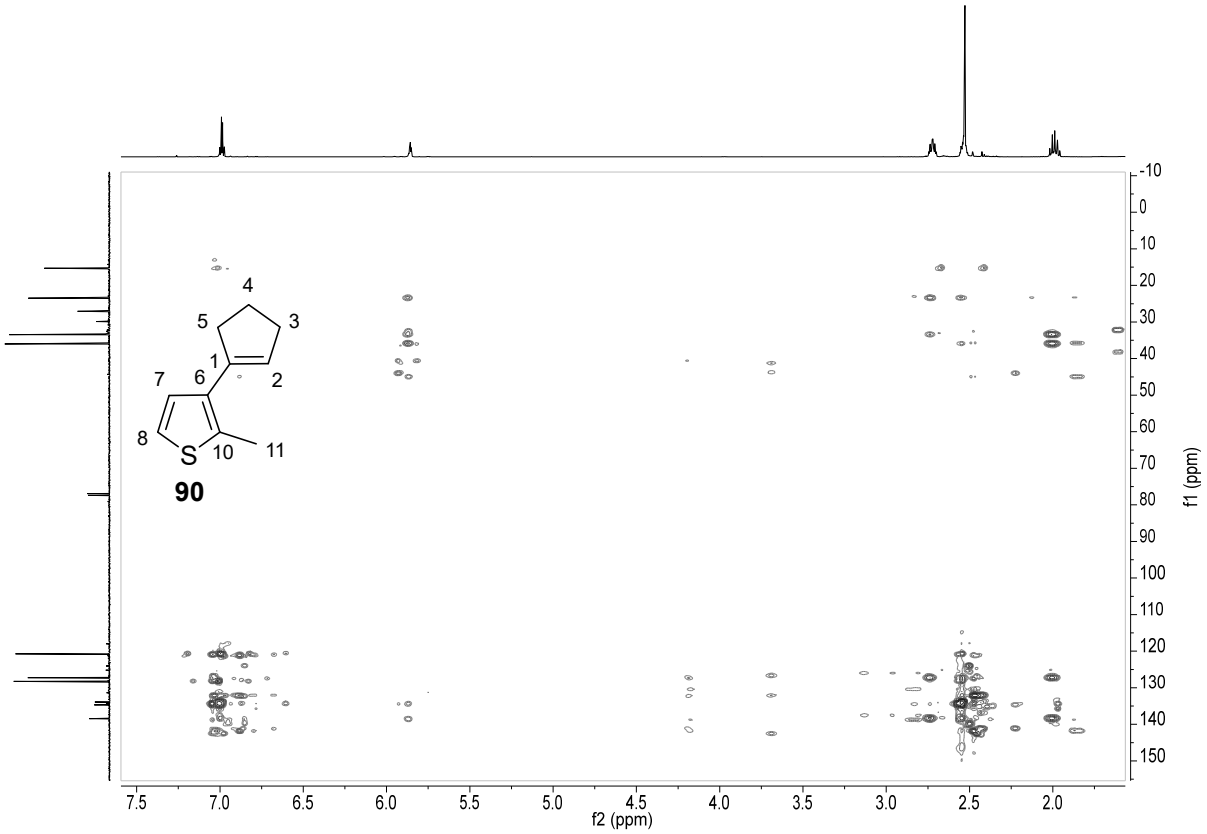


Figure 235. HMBC-NMR spectrum of cyclopentene-thiophene adduct **90**.

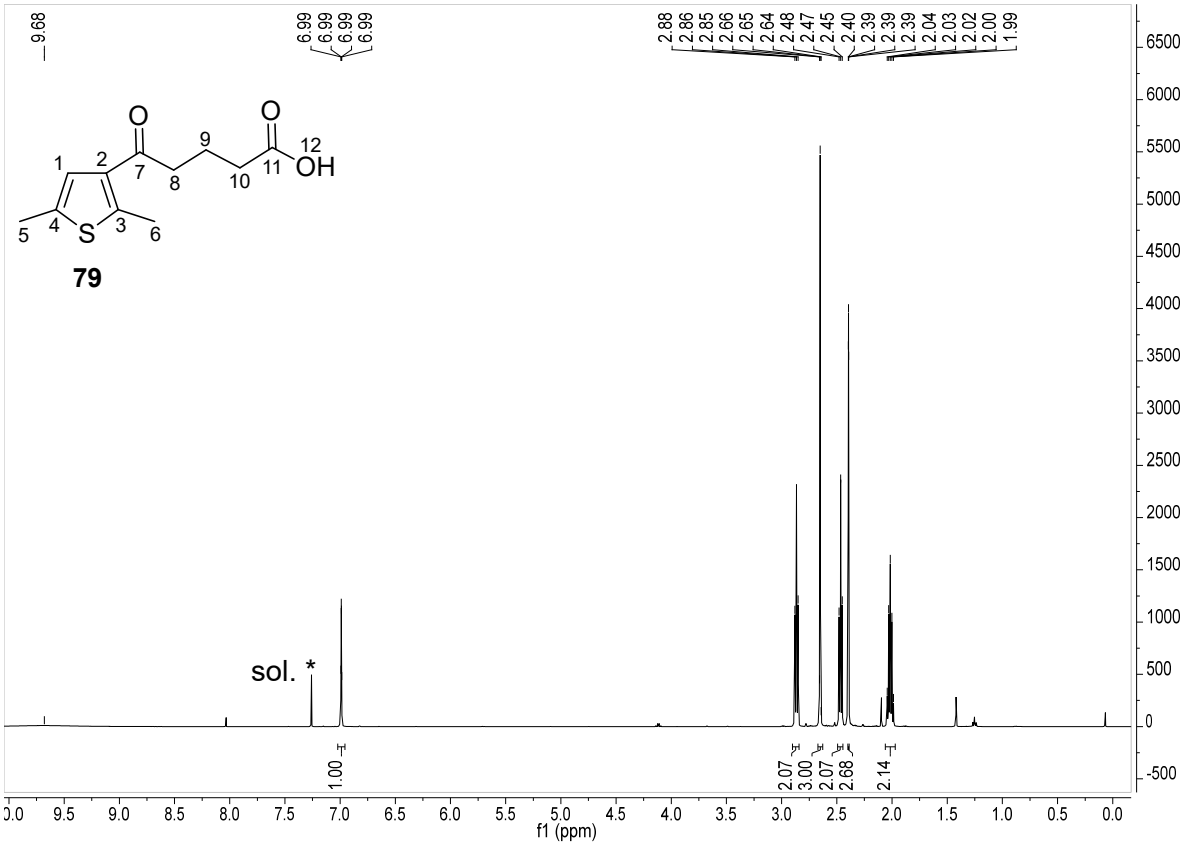


Figure 236. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of pentanoic acid derivative **79**.

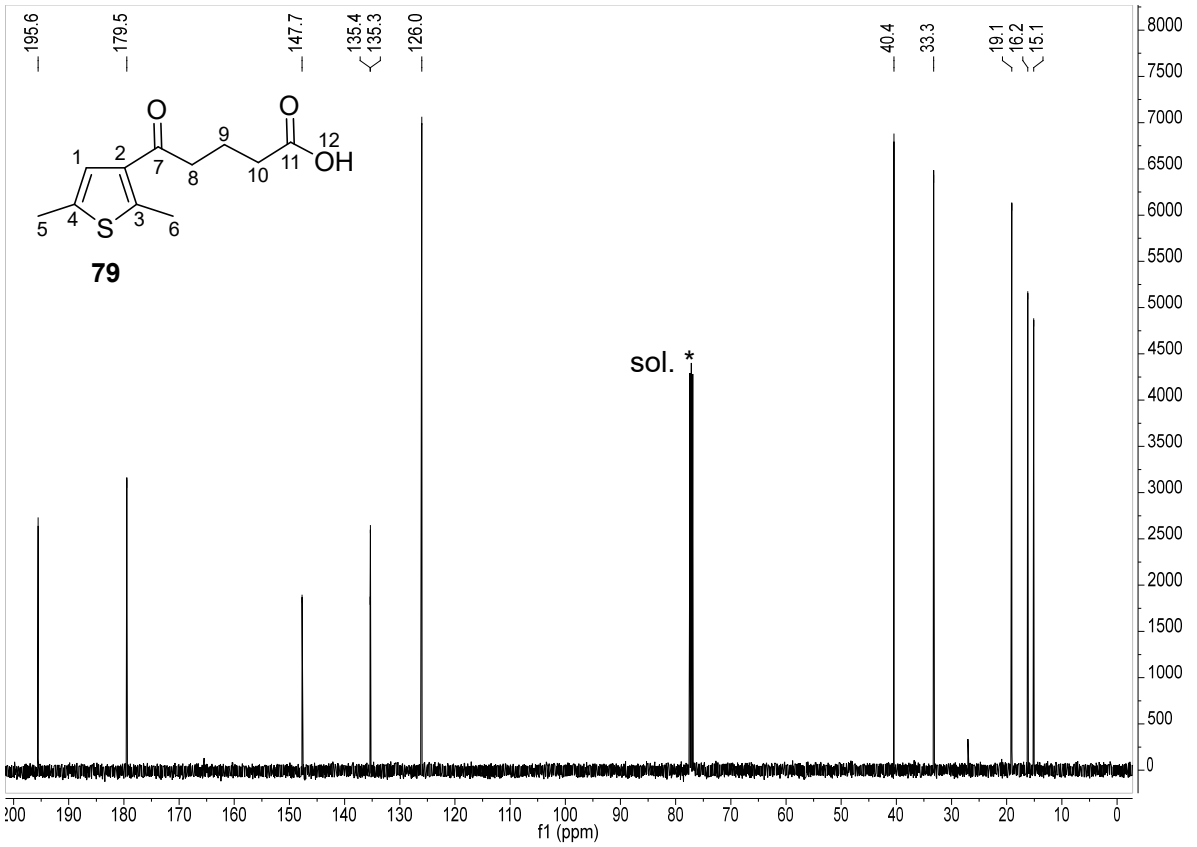


Figure 237. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of pentanoic acid derivative **79**.

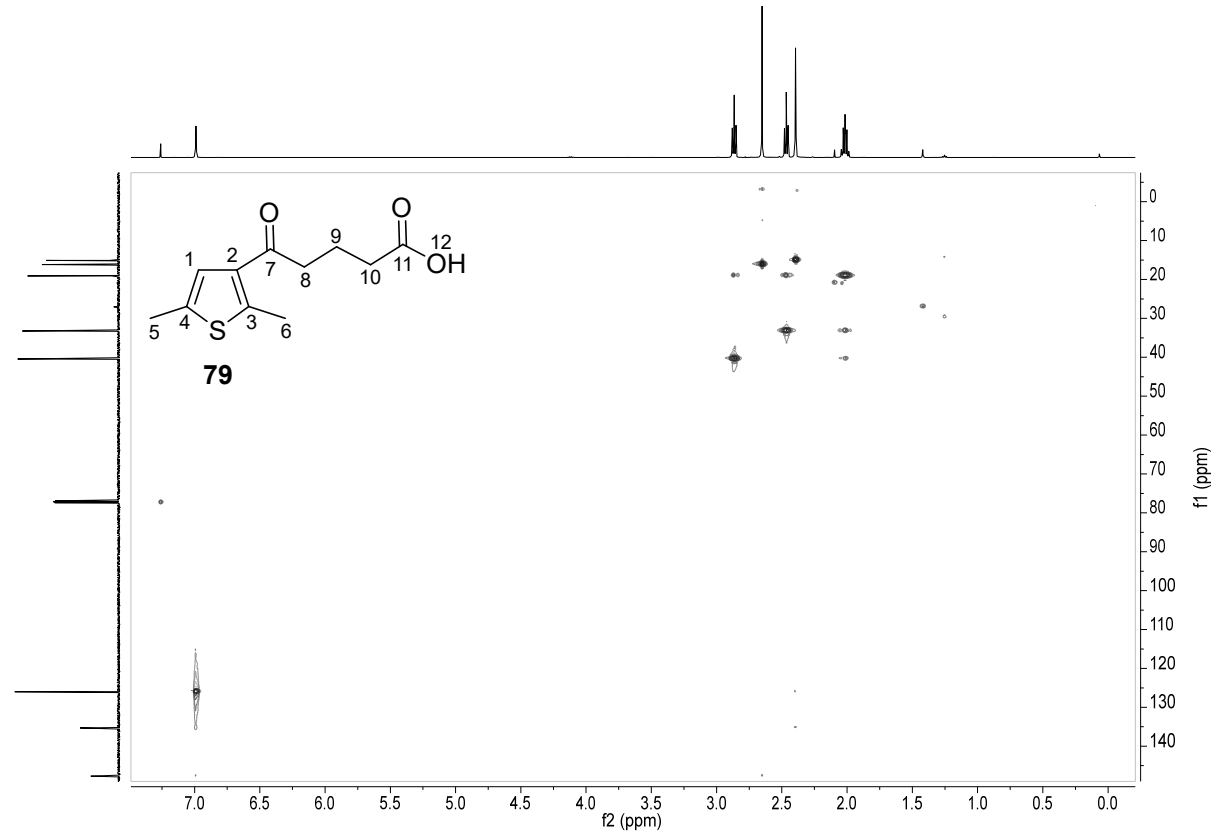


Figure 238. HSCQ-NMR spectrum of pentanoic acid derivative **79**.

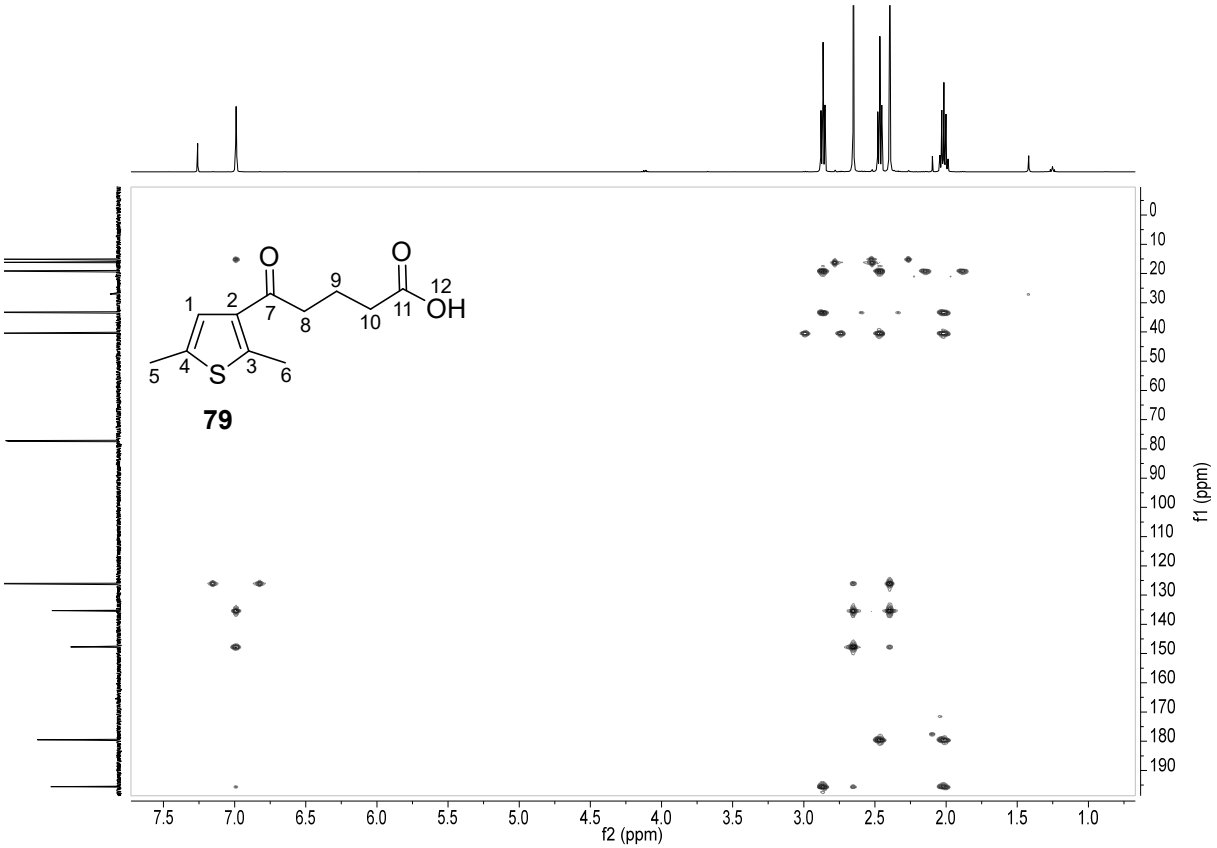


Figure 239. HMBC-NMR spectrum of pentanoic acid derivative **79**.

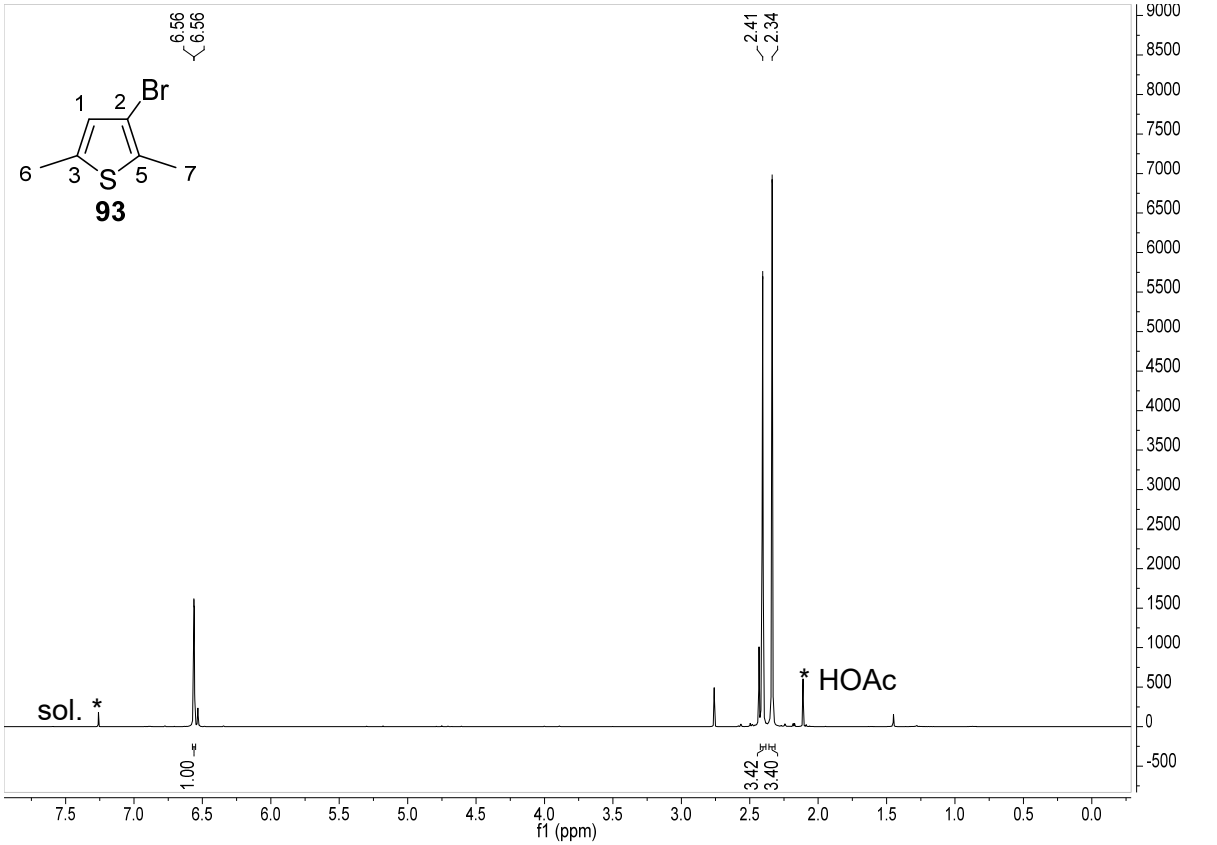


Figure 240. ^{1}H-NMR ($CDCl_3$, 400 MHz) of thiophene **93**.

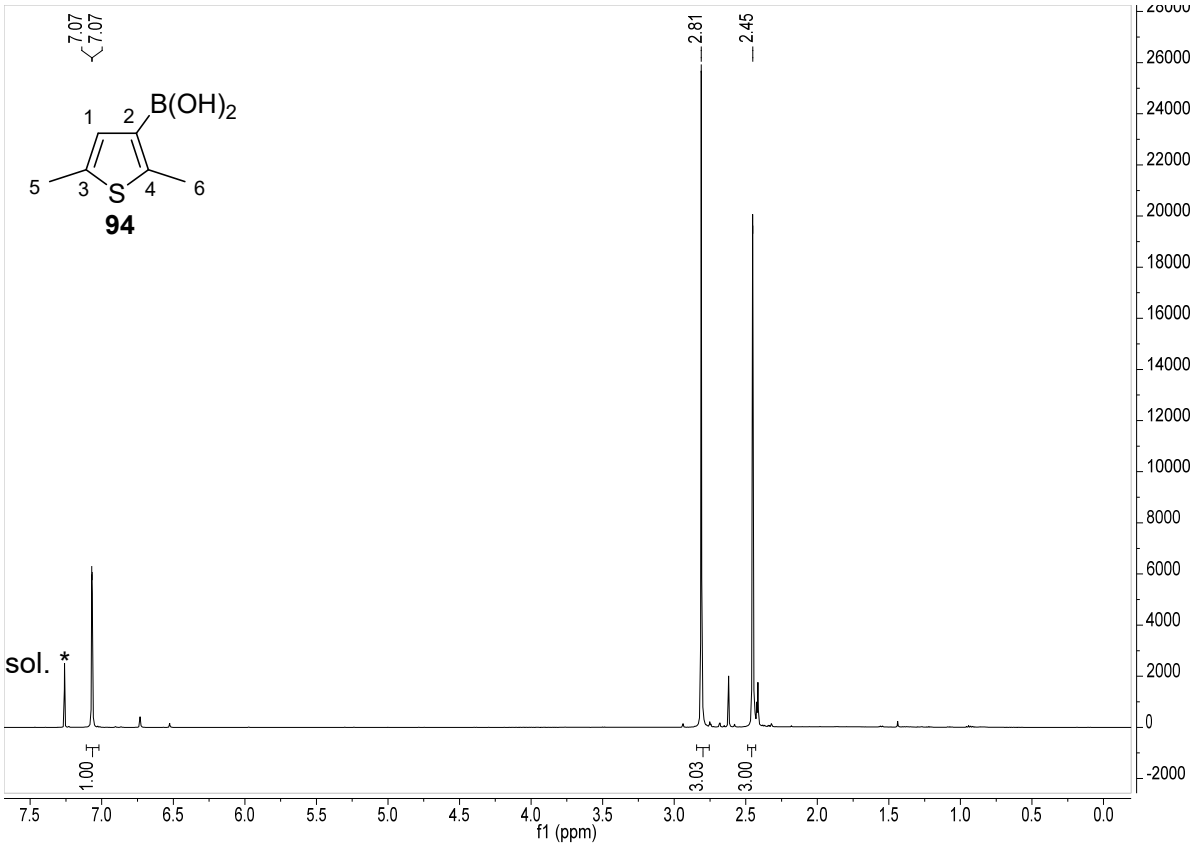


Figure 241. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of thienyl boronic acid **94**.

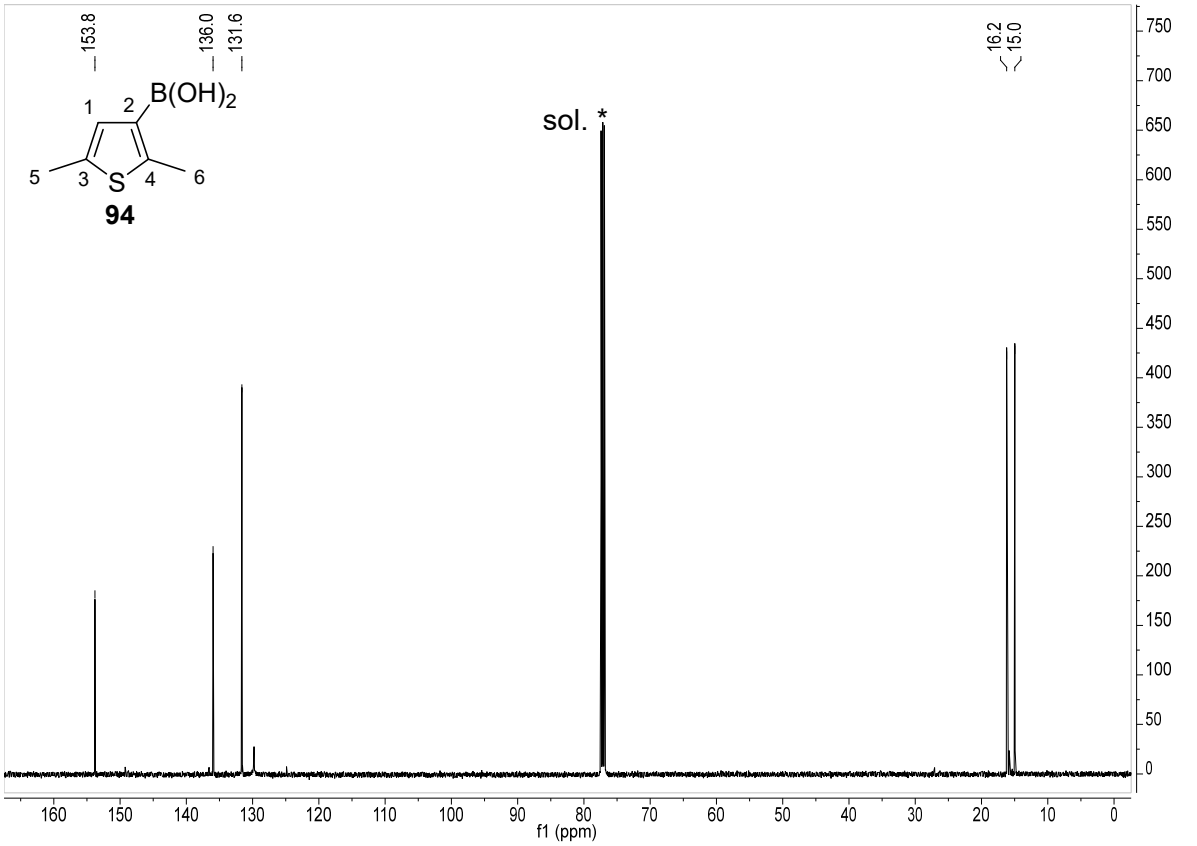


Figure 242. ^{13}C-NMR spectrum ($CDCl_3$, 126 MHz) of thienyl boronic acid **94**.

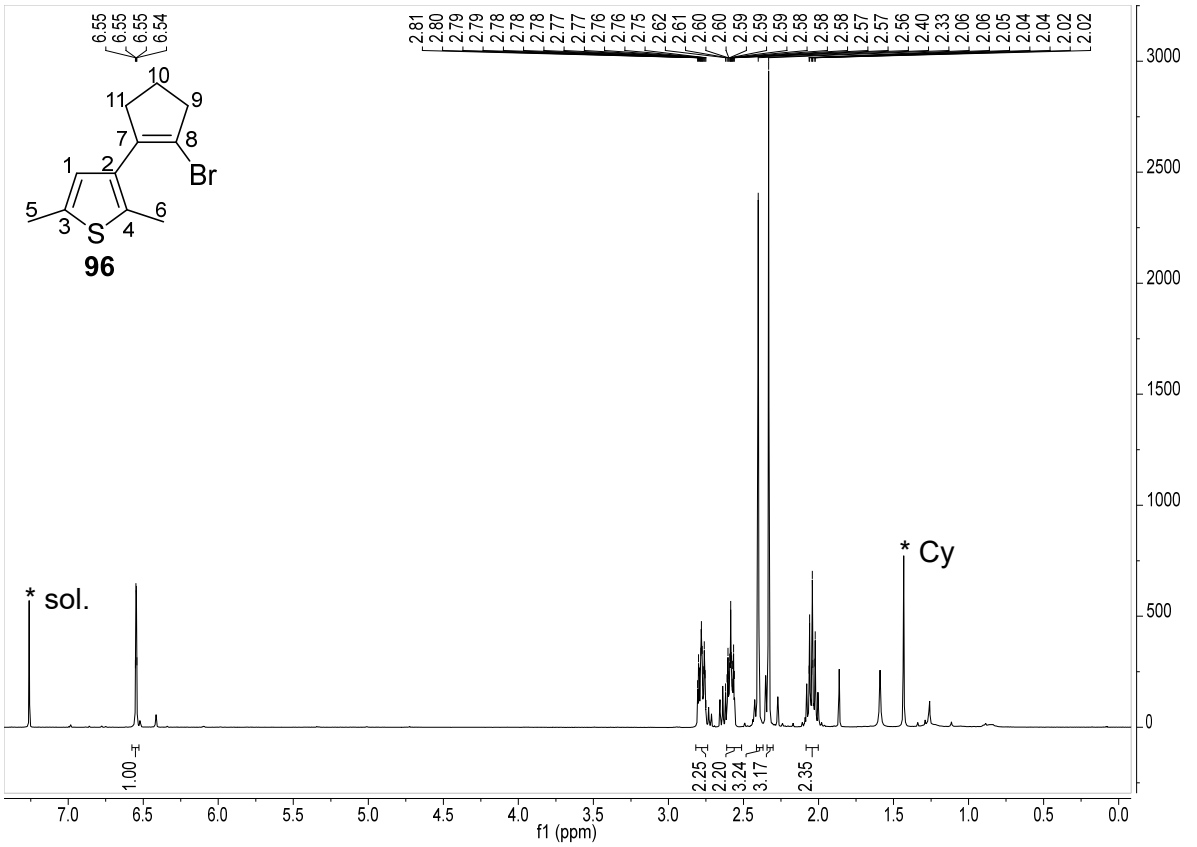


Figure 243. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **96**.

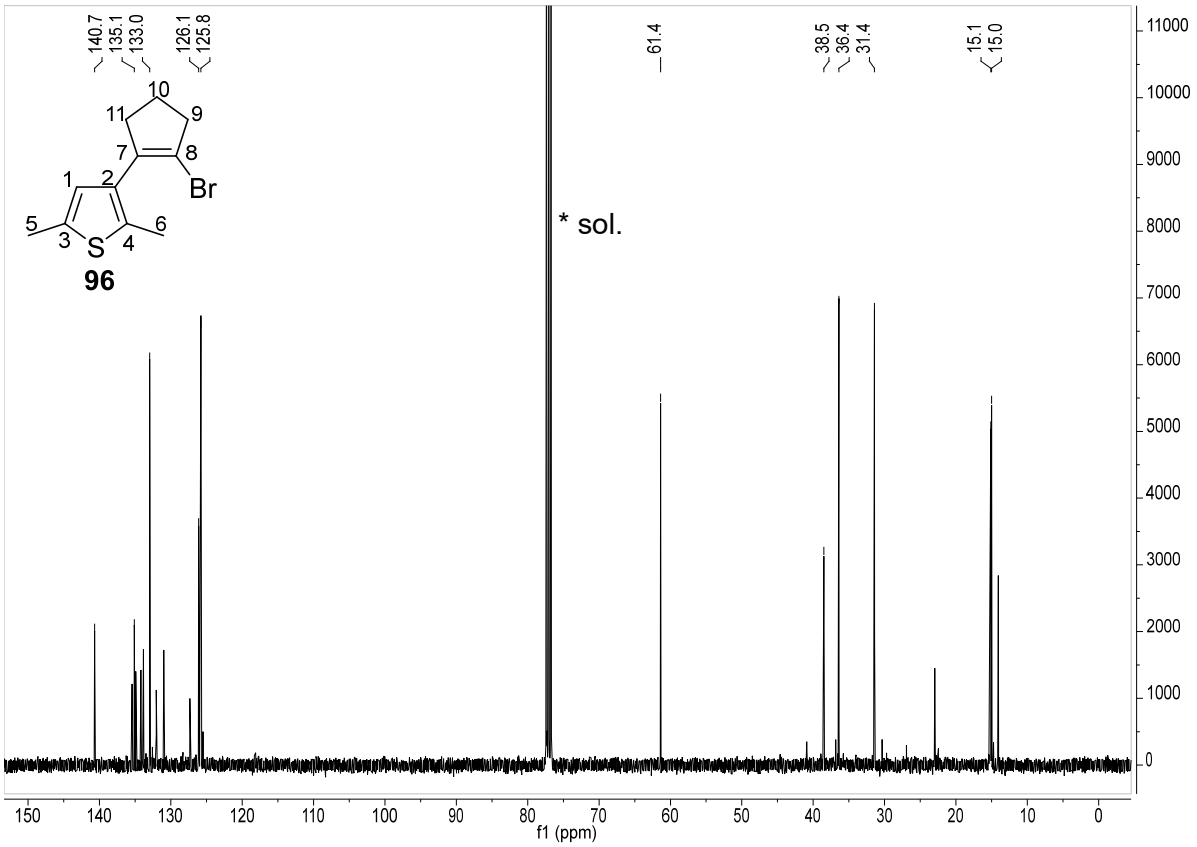


Figure 244. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of compound **96**.

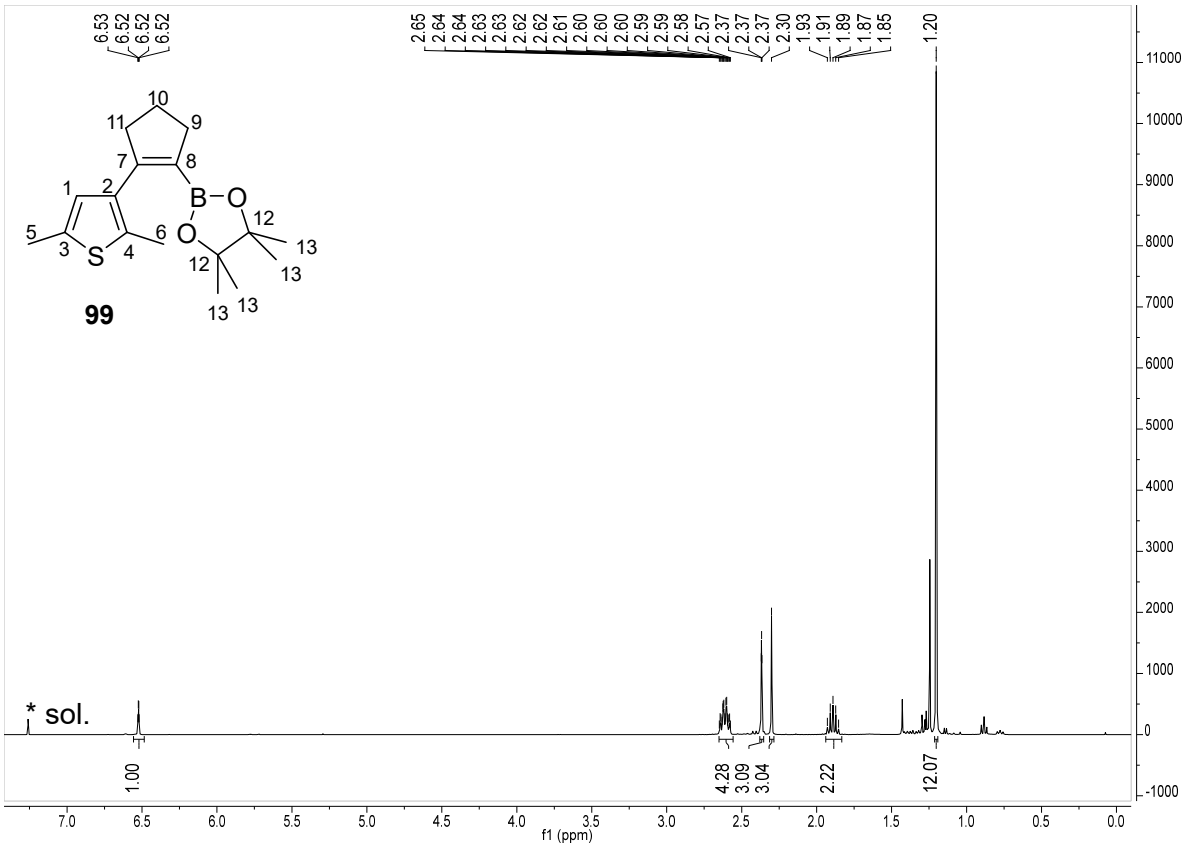


Figure 245. ^{1}H-NMR spectrum ($CDCl_3$, 400 MHz) of compound **99**.

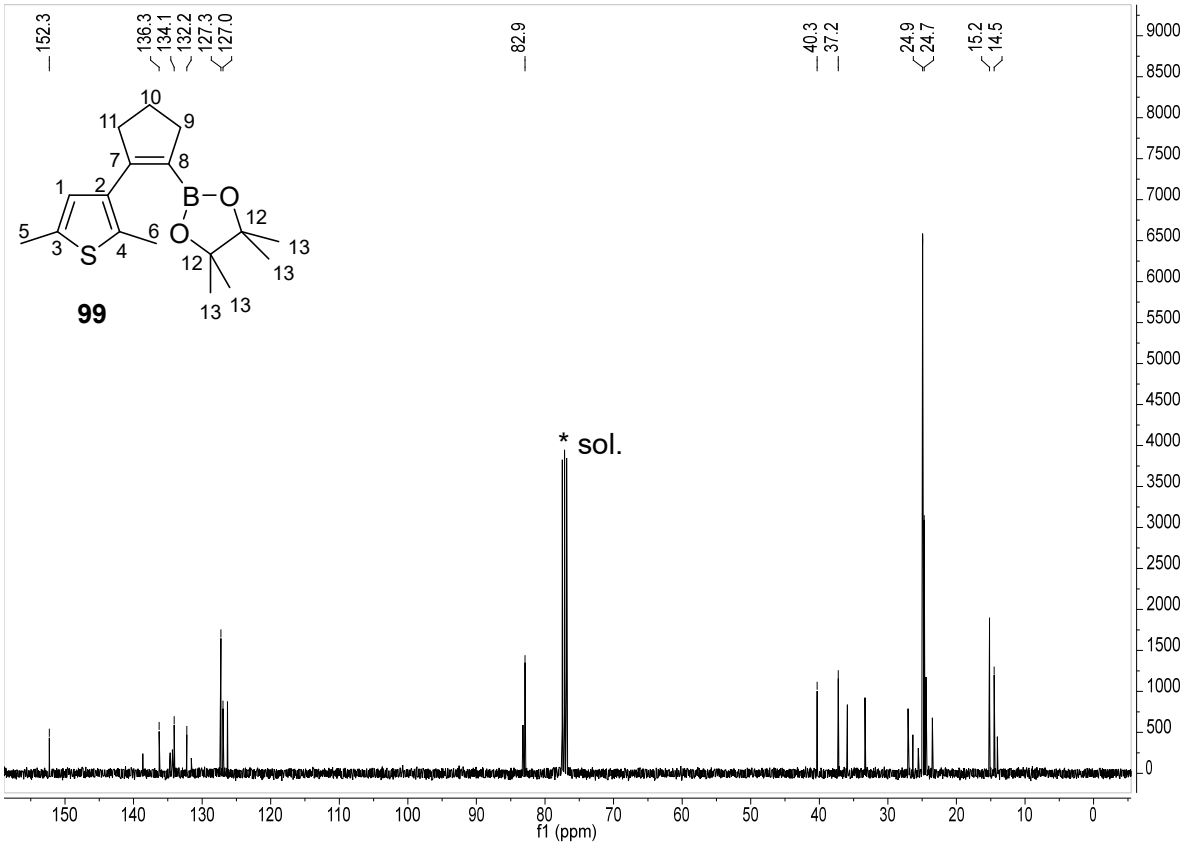


Figure 246. ^{13}C-NMR spectrum ($CDCl_3$, 101 MHz) of compound **99**.

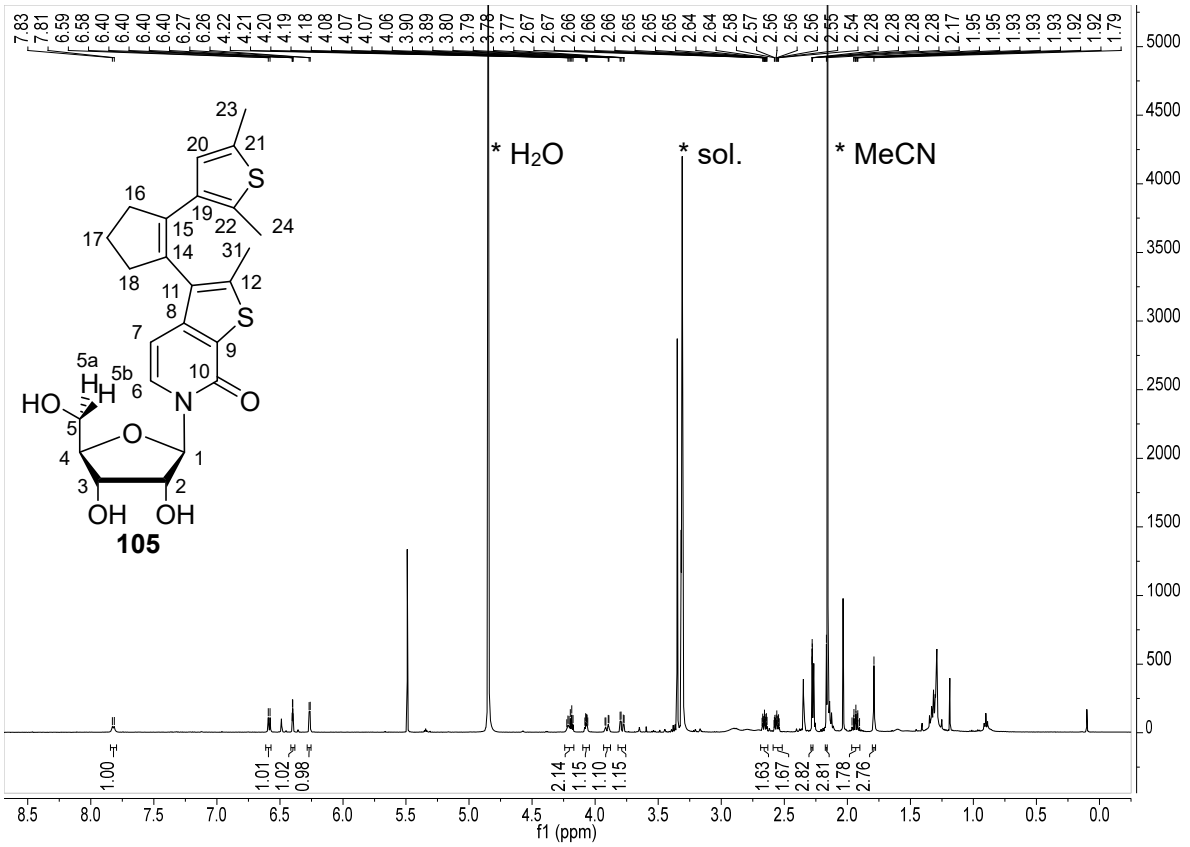


Figure 247. ^{1}H-NMR spectrum (CD_3OD, 500 MHz) of compound **105**.

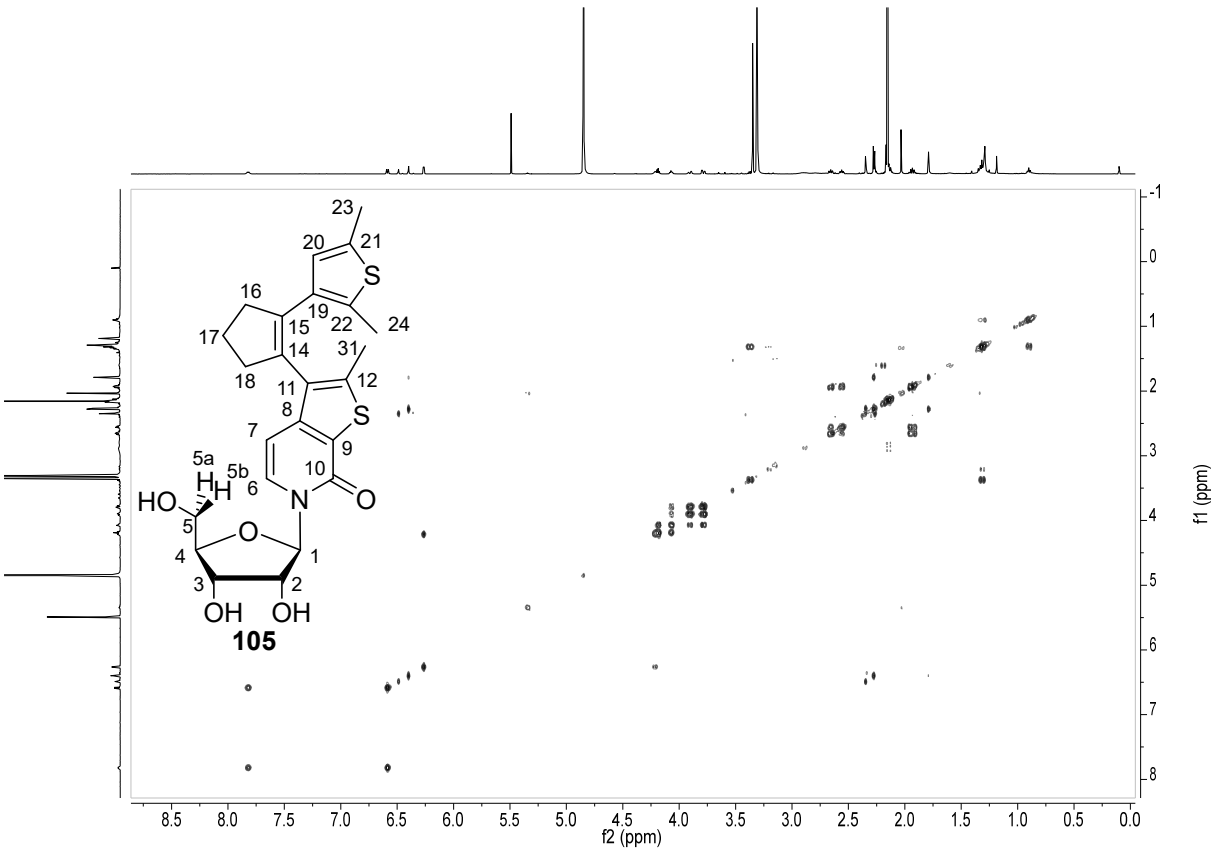


Figure 248. H,H-COSY-NMR spectrum of compound **105**.

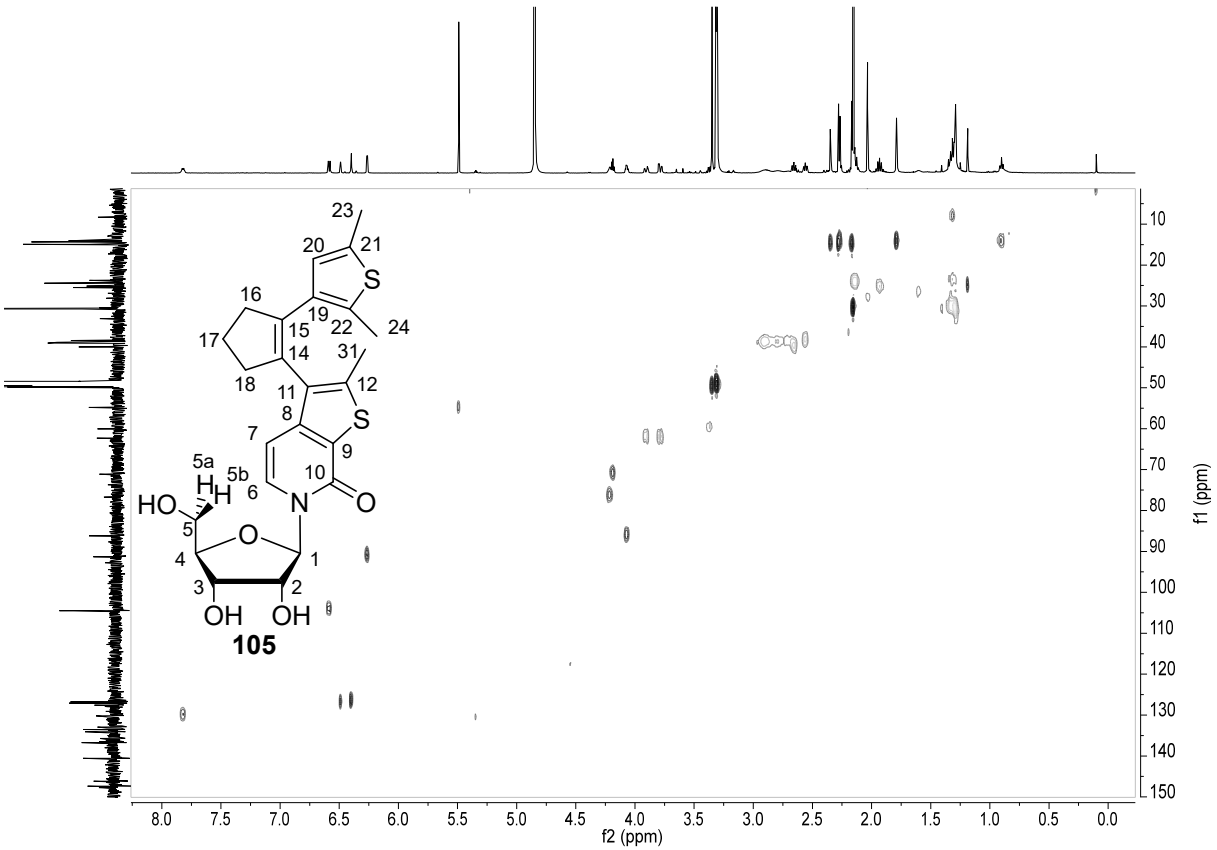


Figure 249. HSQC-NMR spectrum (CD_3OD) of compound **105**.

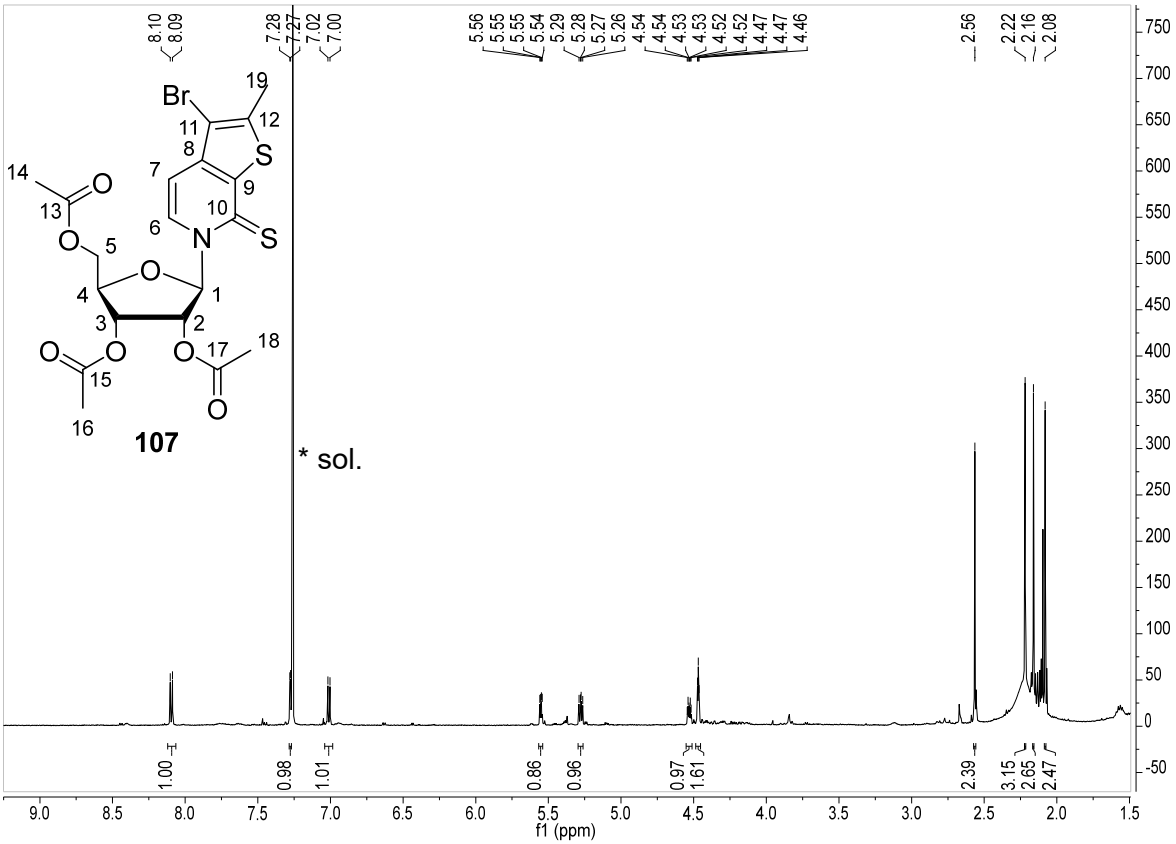


Figure 250. ^{1}H-NMR spectrum ($CDCl_3$, 500 MHz) of compound **107**.

7.2 EPR spectra

Here, raw PELDOR time traces corresponding to the respective data presented in section 3.2.2 are shown and include intermolecular background correction functions as well as polynomial fits.

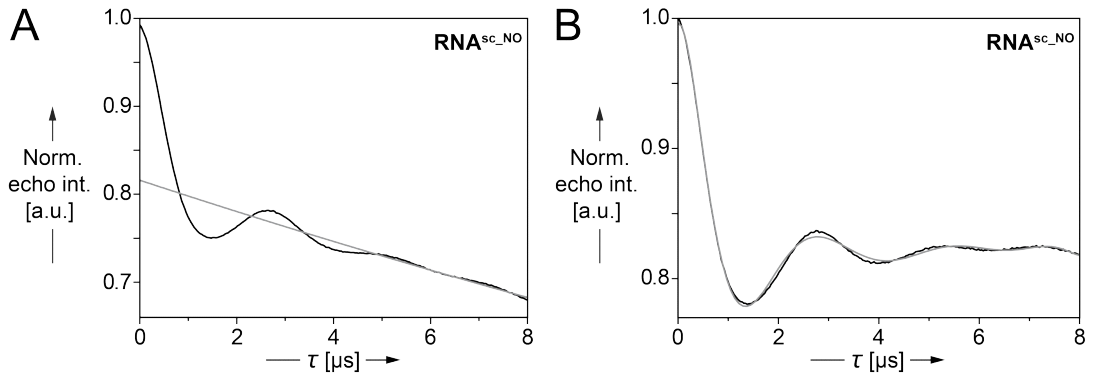


Figure 251. Echo time traces of the **RNA^{sc_NO}** PELDOR analysis shown in Figure 80; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

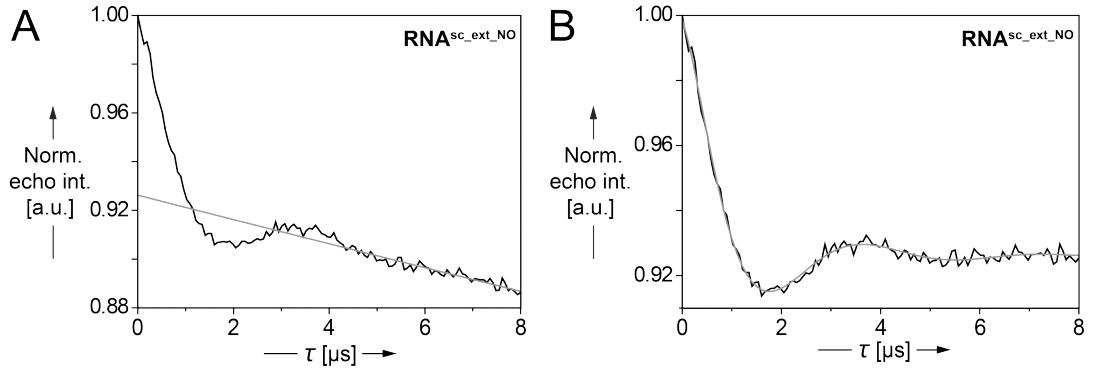


Figure 252. Echo time traces of the **RNA^{sc_ext_NO}** PELDOR analysis shown in Figure 81; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

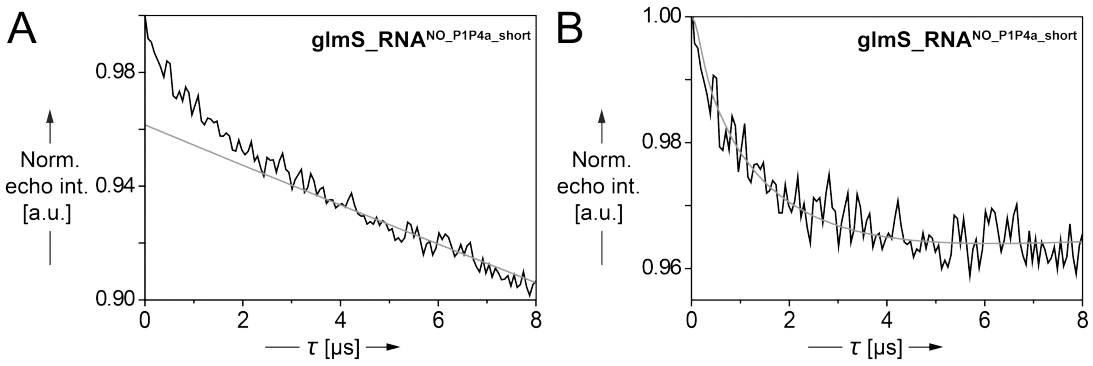


Figure 253. Echo time traces of the **glmS_RNA^{NO_P1P4a_short}** PELDOR analysis shown in Figure 93; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

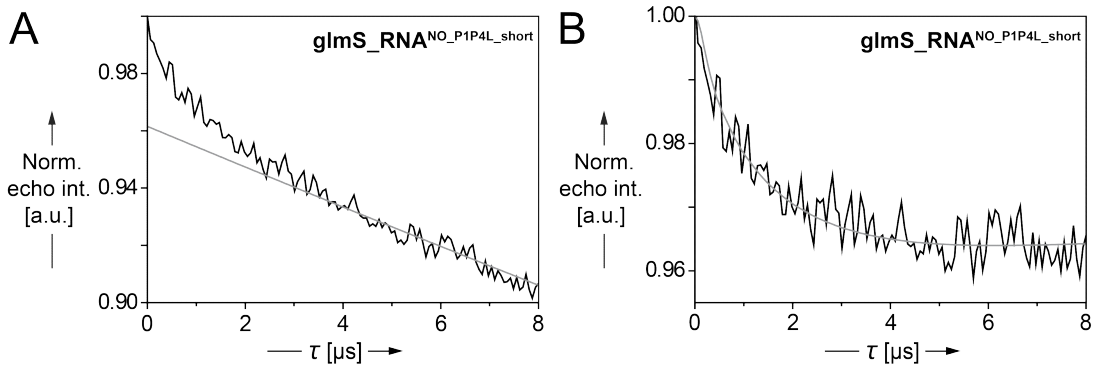


Figure 254. Echo time traces of the **glmS_RNA^{NO_P1P4L_short}** PELDOR analysis shown in Figure 93; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

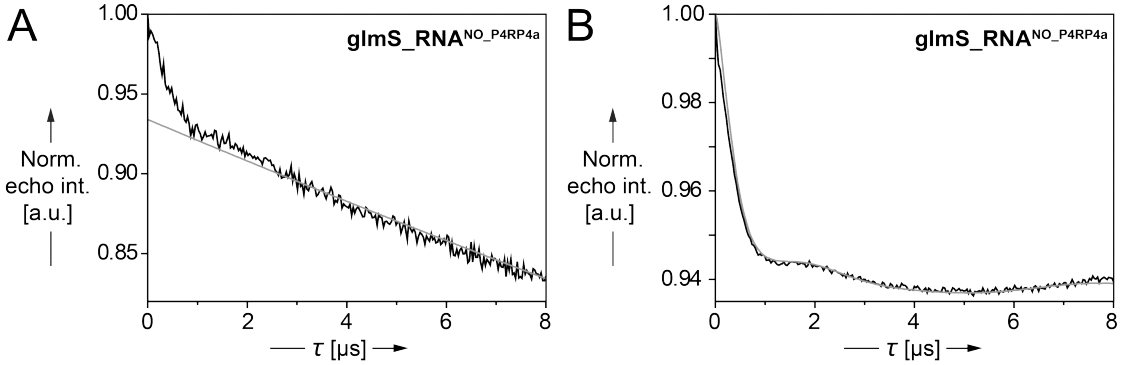


Figure 255. Echo time traces of the **glmS_RNA^{NO_P4RP4a}** PELDOR analysis shown in Figure 96; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

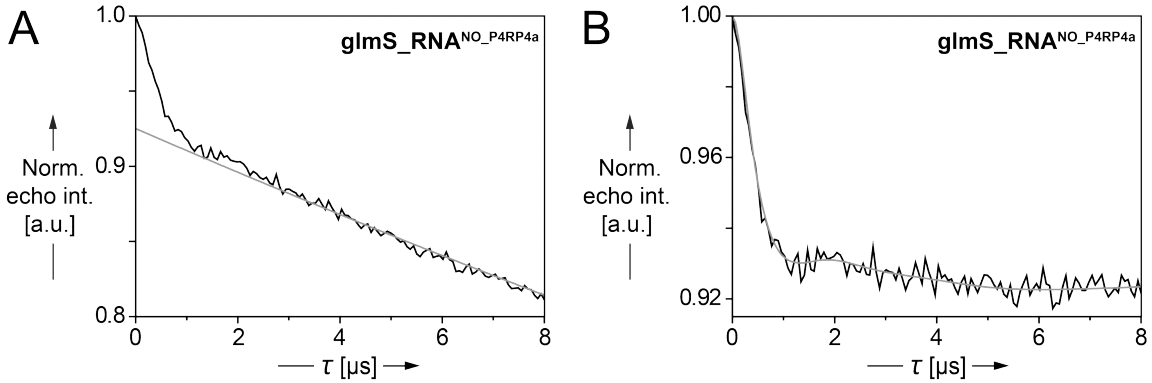


Figure 256. Echo time traces of the **glmS_RNA^{NO_P4RP4a}** PELDOR analysis shown in Figure 103; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

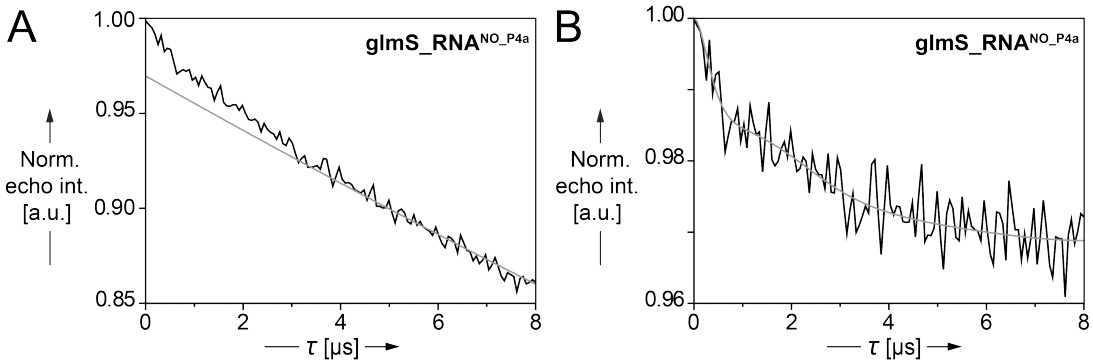


Figure 257. Echo time traces of the **glmS_RNA^{NO_P4a}** PELDOR analysis shown in Figure 104; A. Raw data plot (black) and linear background function (gray); B. Background-corrected data (black) and fitted curve (gray).

7.3 MS spectra

In the following, MS and HR MS spectra necessary to specify the known and novel compounds described in section 5.2.7 are presented. Main peaks corresponding to the species of interest are denoted.

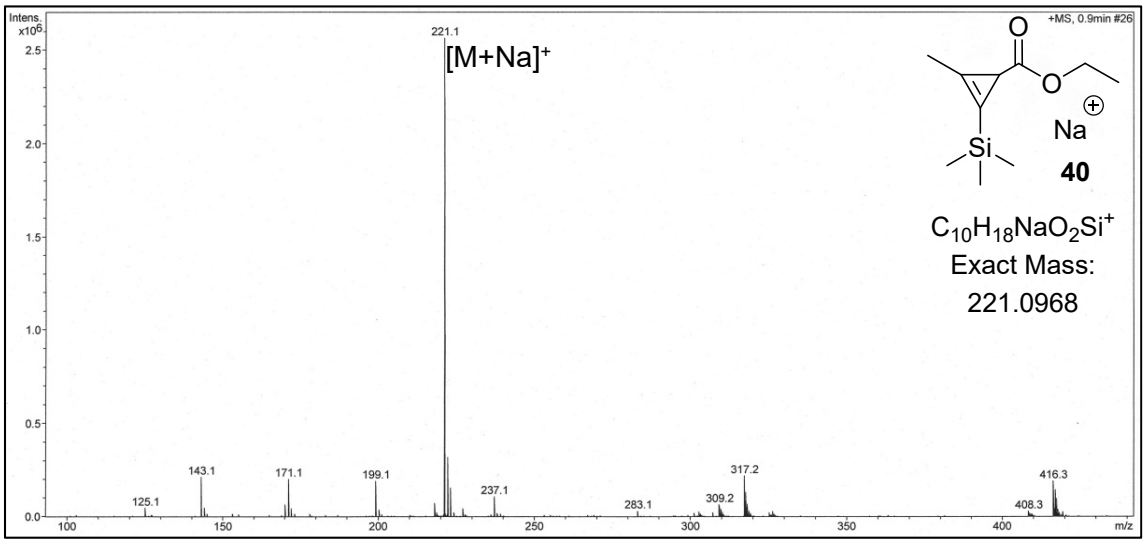


Figure 258. ESI$^+$ MS spectrum of compound **40**.

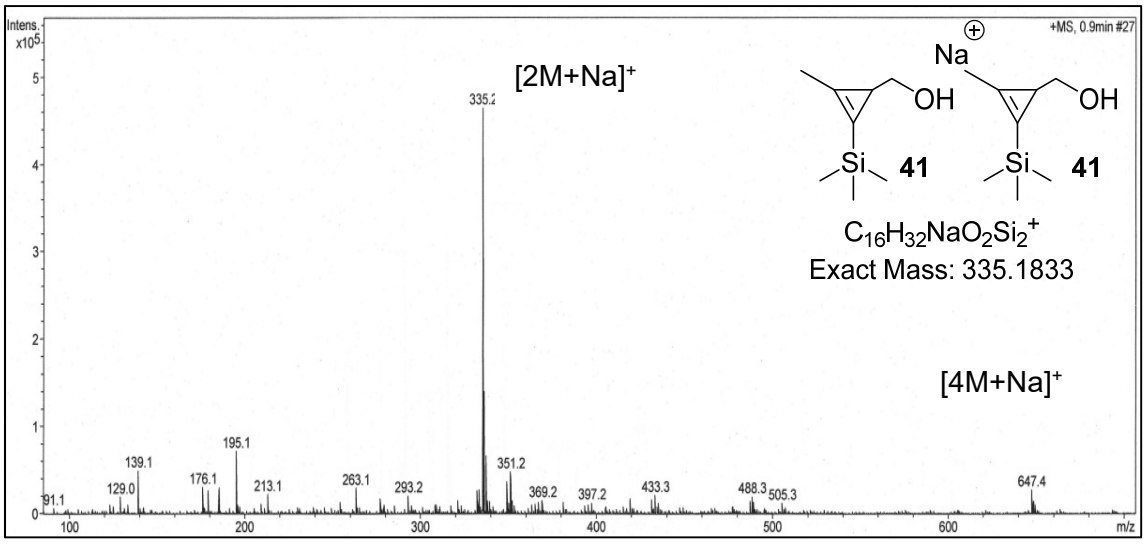


Figure 259. ESI+ MS spectrum of compound **41**.

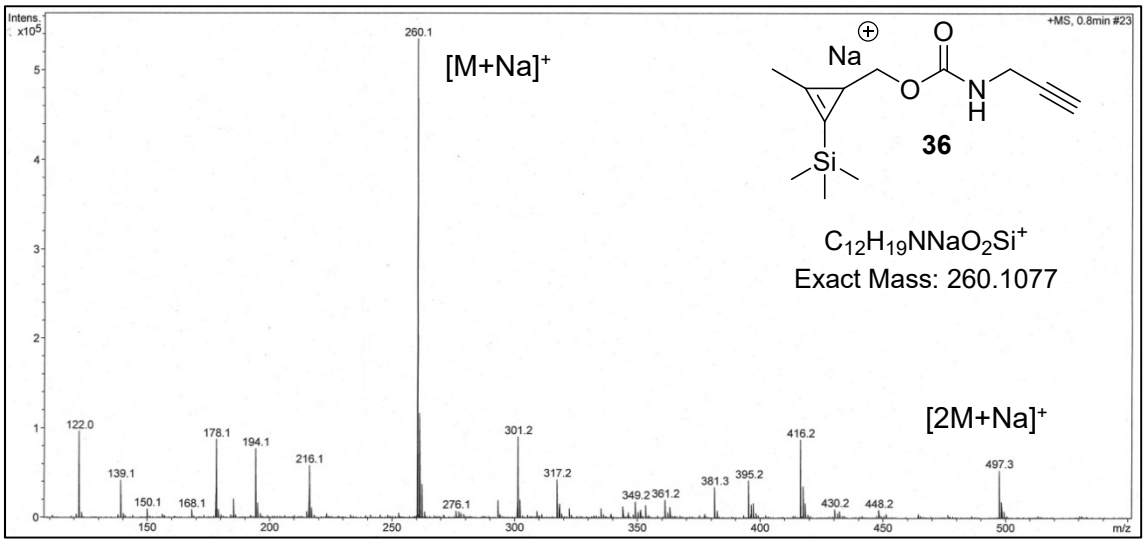


Figure 260. ESI+ MS spectrum of compound **36**.

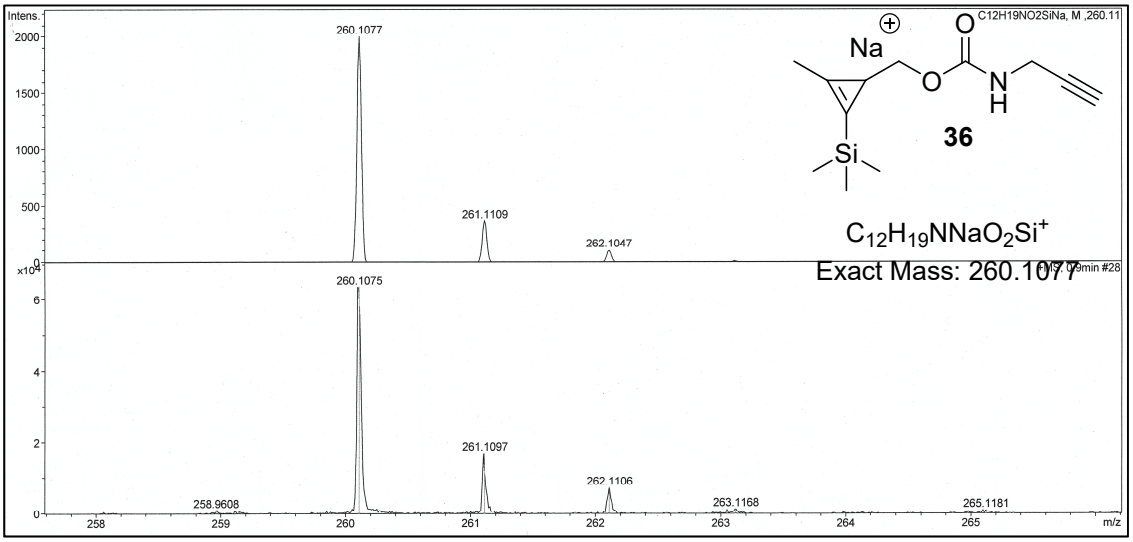


Figure 261. ESI+ HR MS spectrum of compound **36**.

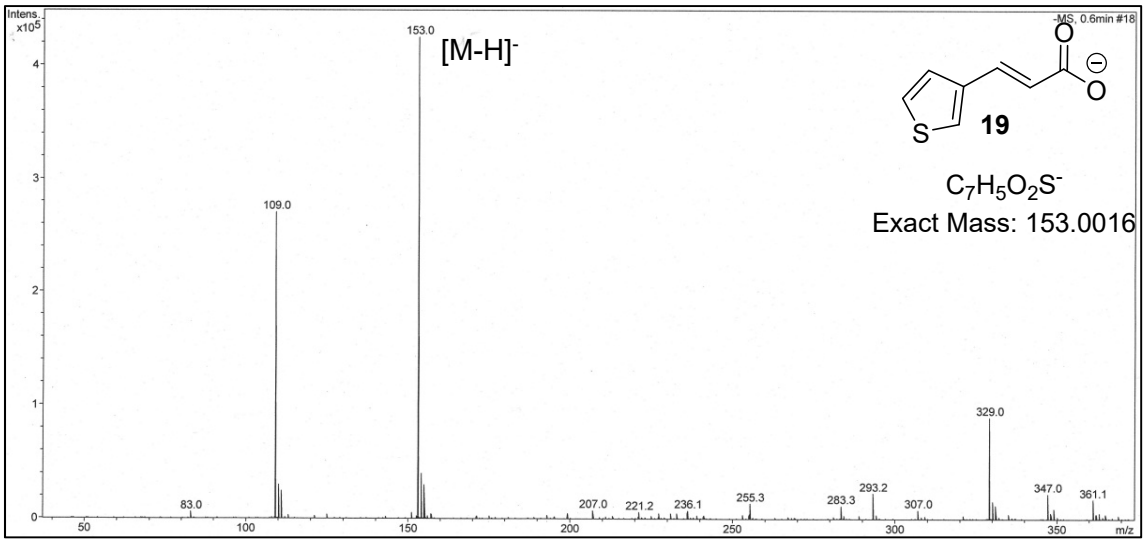


Figure 262. ESI$^-$ MS spectrum of compound **19**.

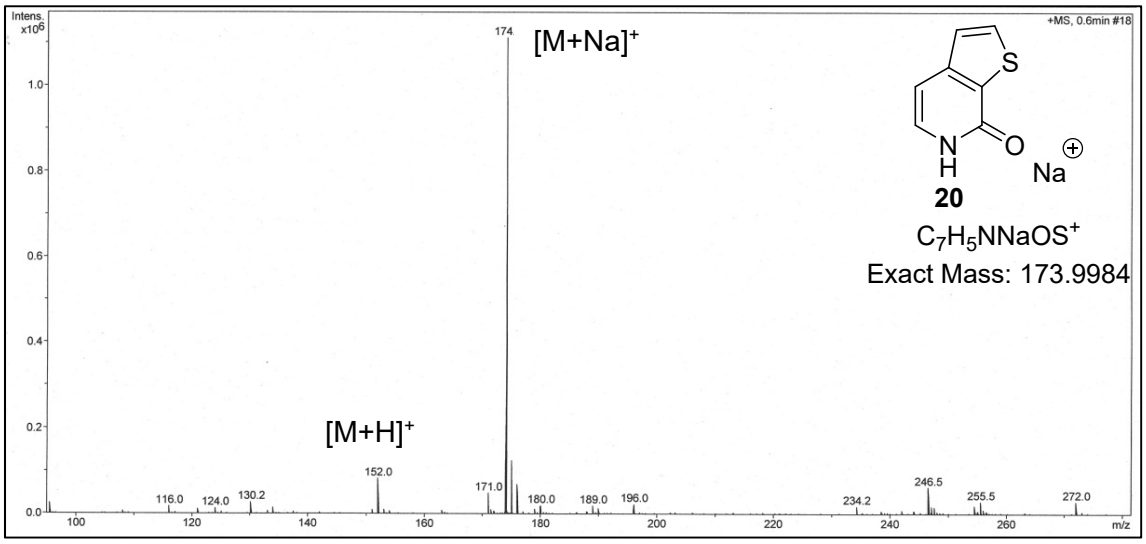


Figure 263. ESI$^+$ MS spectrum of compound **20**.

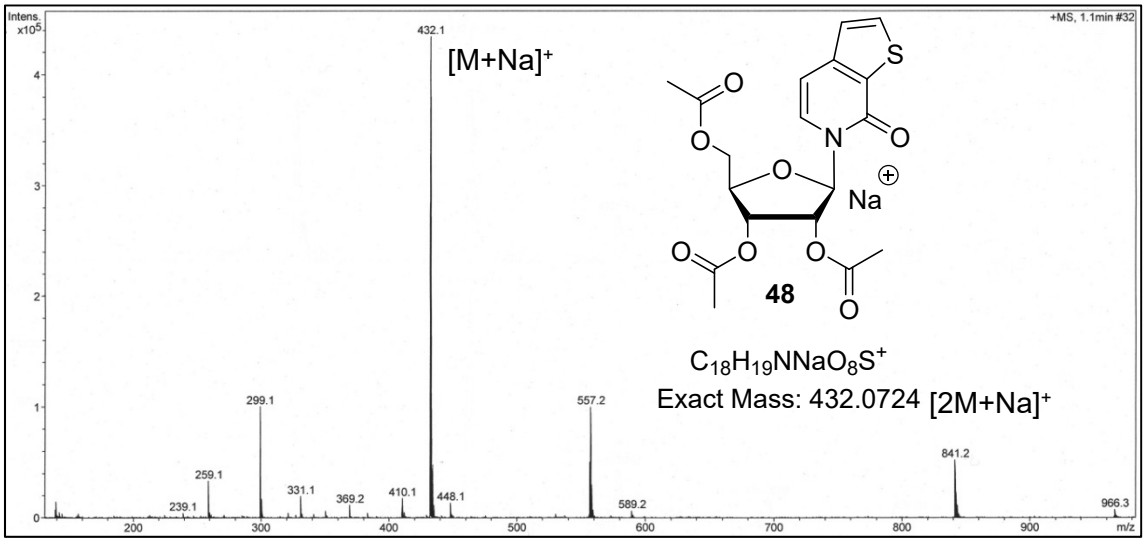


Figure 264. ESI$^+$ MS spectrum of compound **48**.

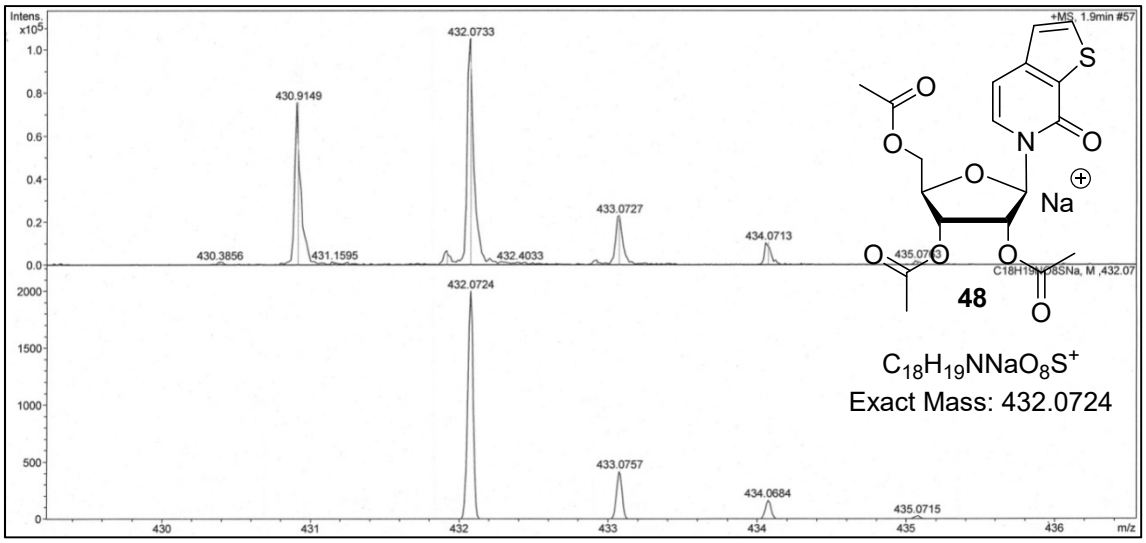


Figure 265. ESI⁺ HR MS spectrum of compound **48**.

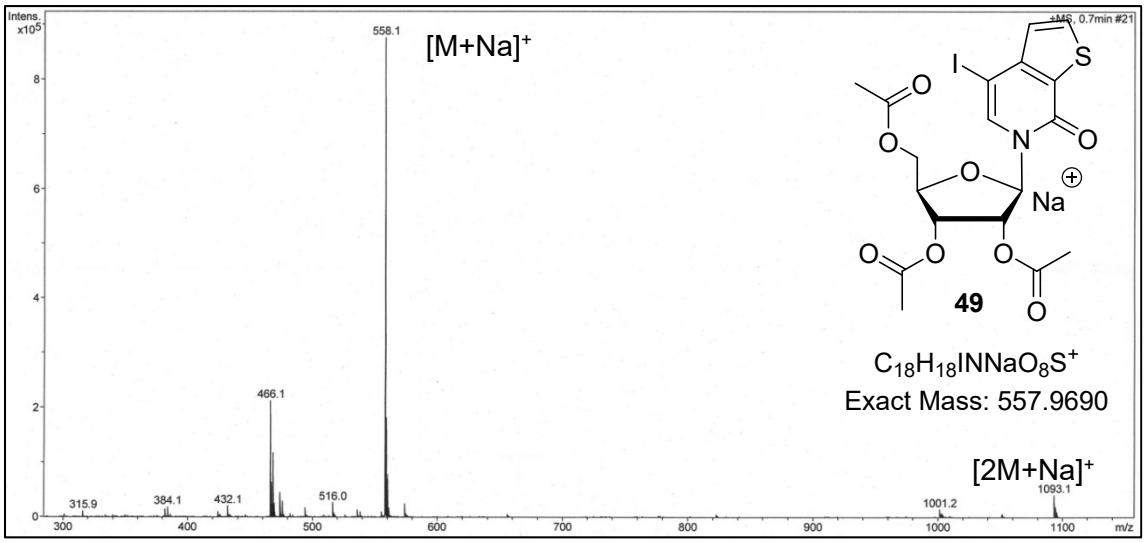


Figure 266. ESI⁺ MS spectrum of compound **49**.

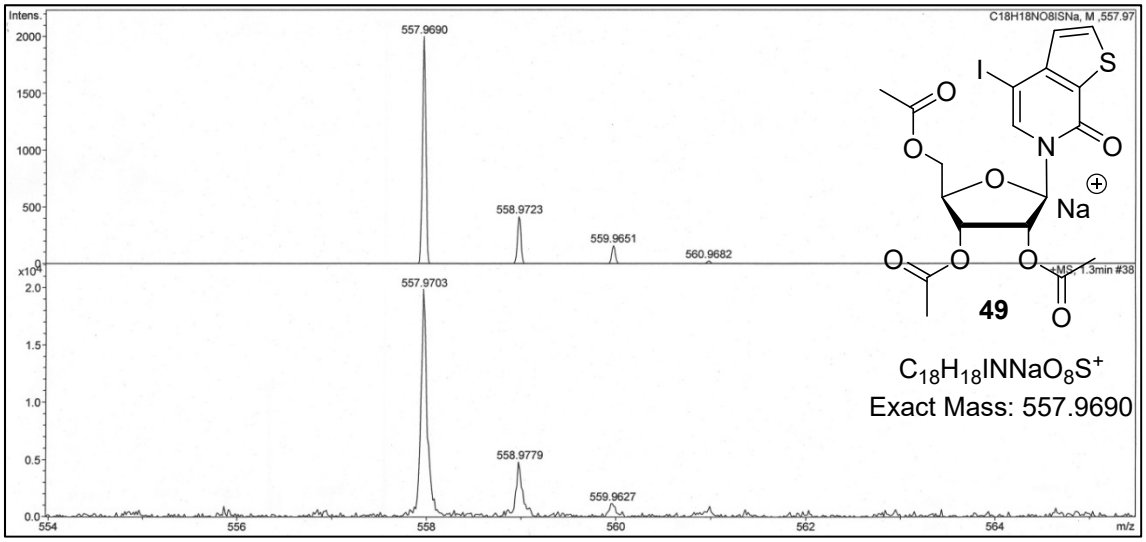


Figure 267. ESI⁺ HR MS spectrum of compound **49**.

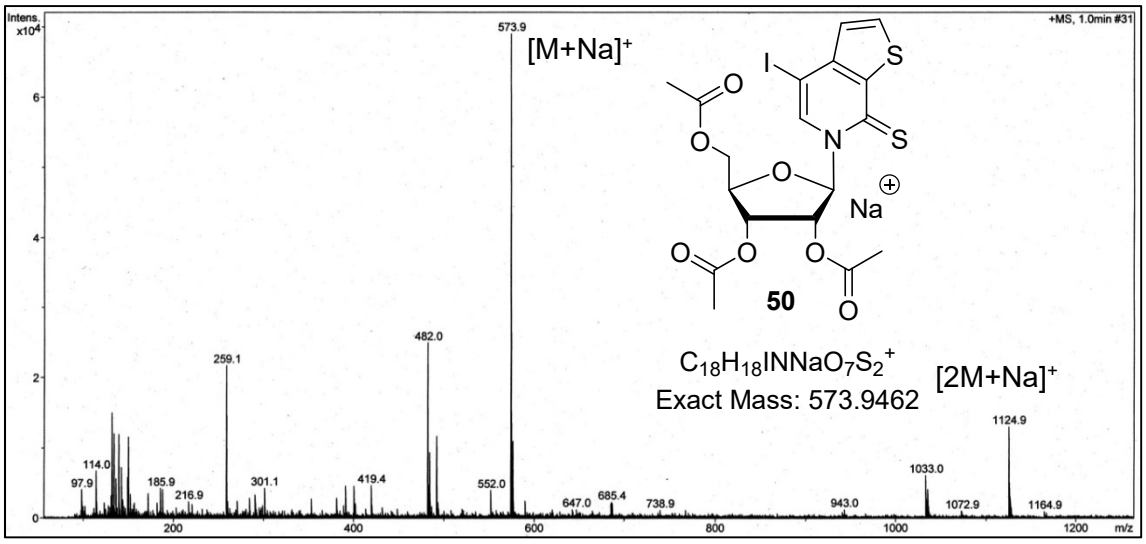


Figure 268. ESI+ MS spectrum of compound **50**.

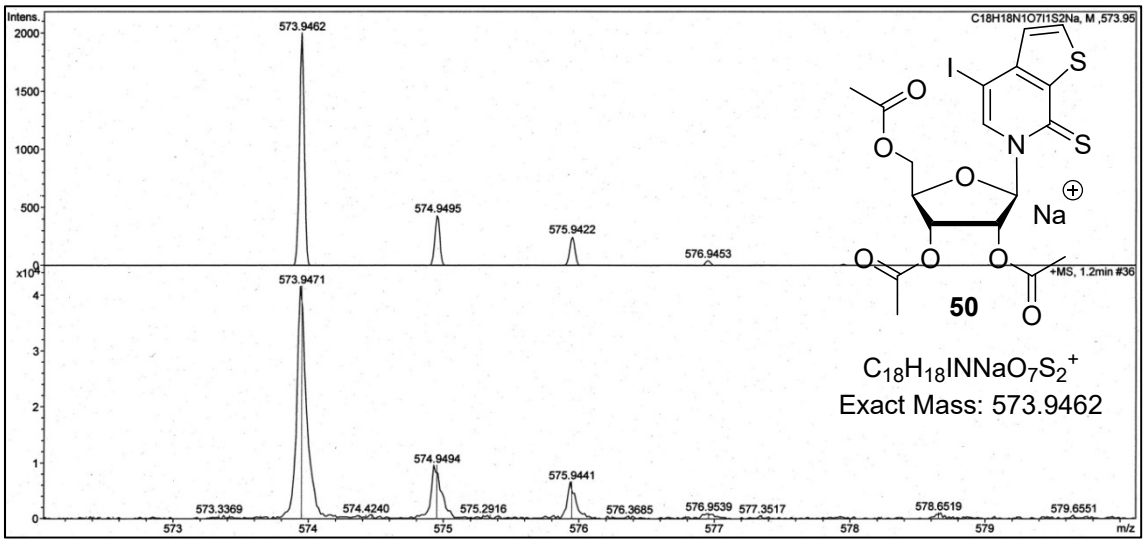


Figure 269. ESI+ HR MS spectrum of compound **50**.

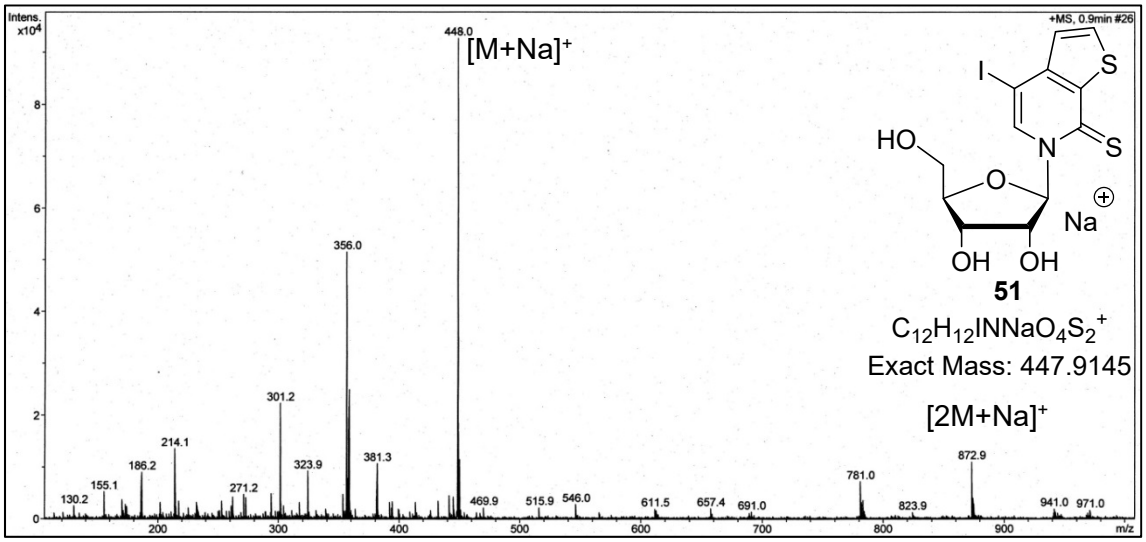


Figure 270. ESI+ MS spectrum of compound **51**.

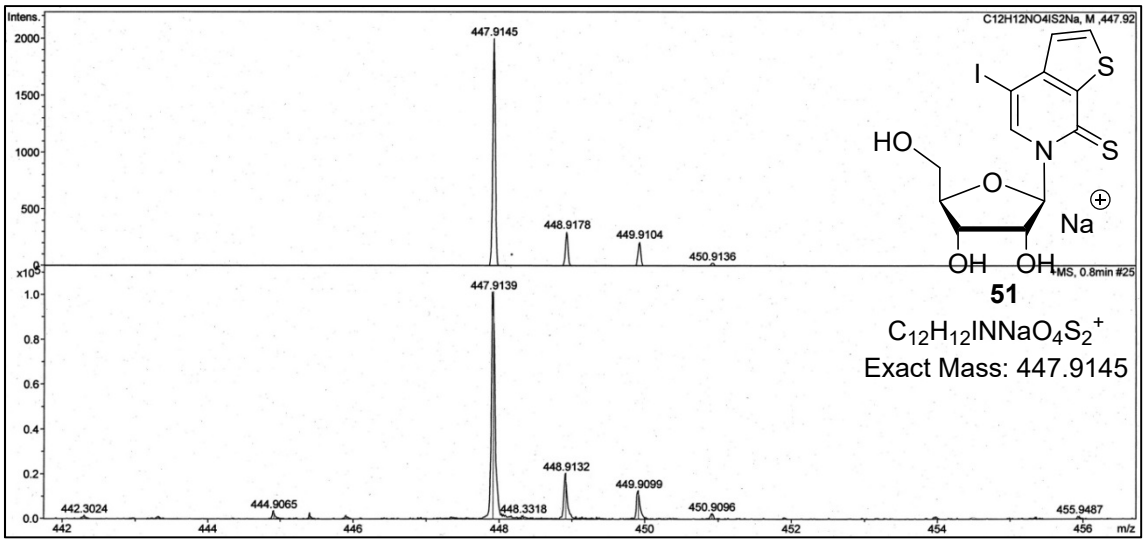


Figure 271. ESI⁺ HR MS spectrum of compound **51**.

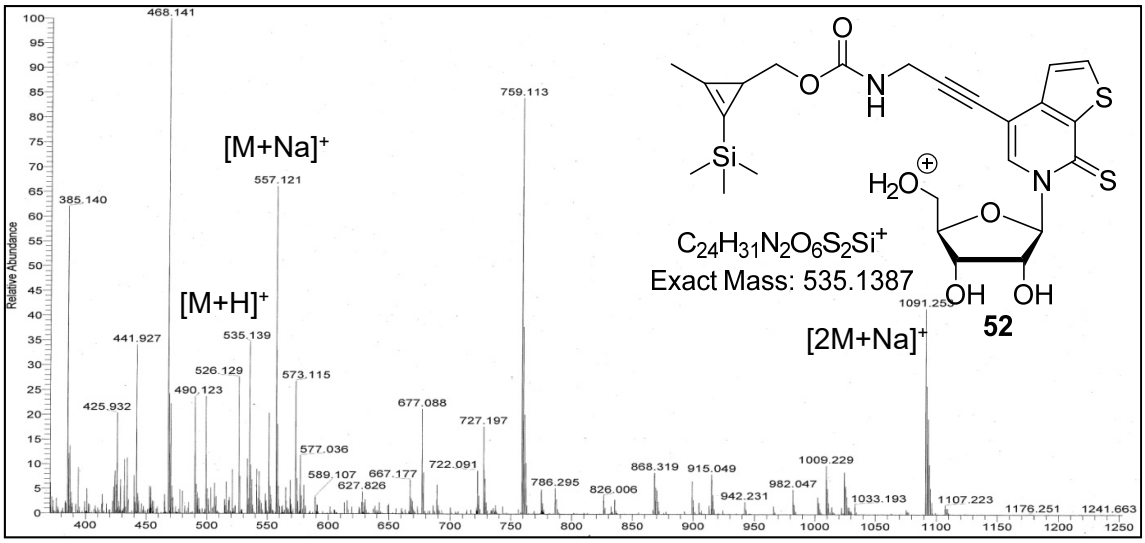


Figure 272. ESI⁺ MS spectrum of compound **52**.

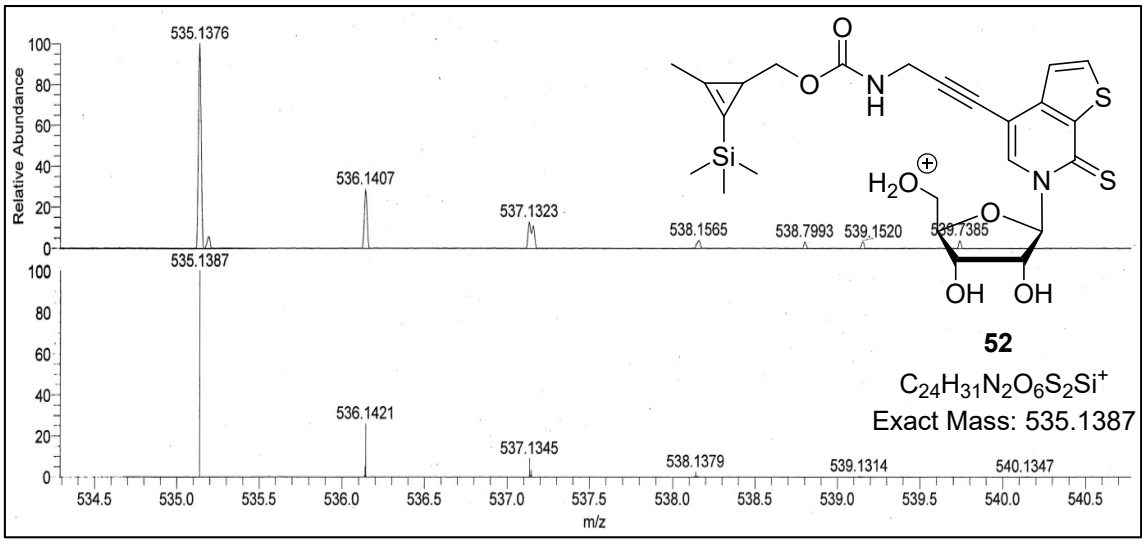


Figure 273. ESI⁺ HR MS spectrum of compound **52**.

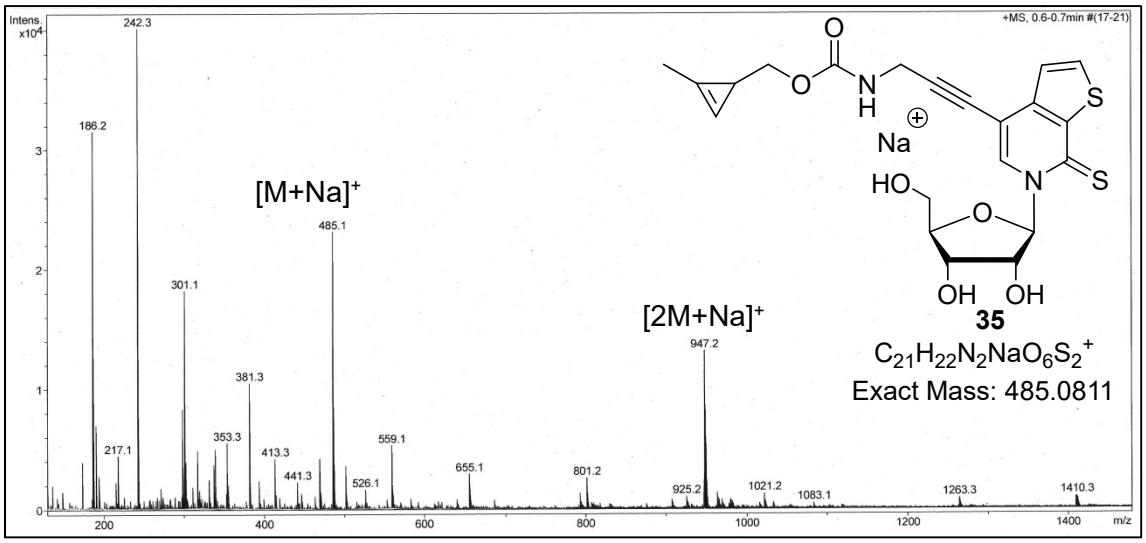


Figure 274. ESI$^+$ MS spectrum of compound **35**.

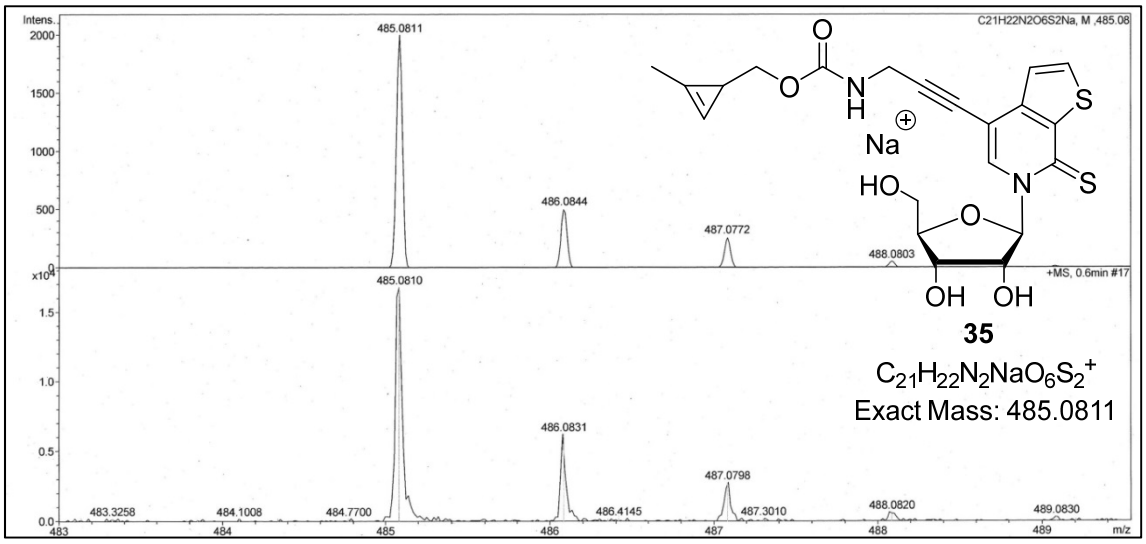


Figure 275. ESI$^+$ HR MS spectrum of compound **35**.

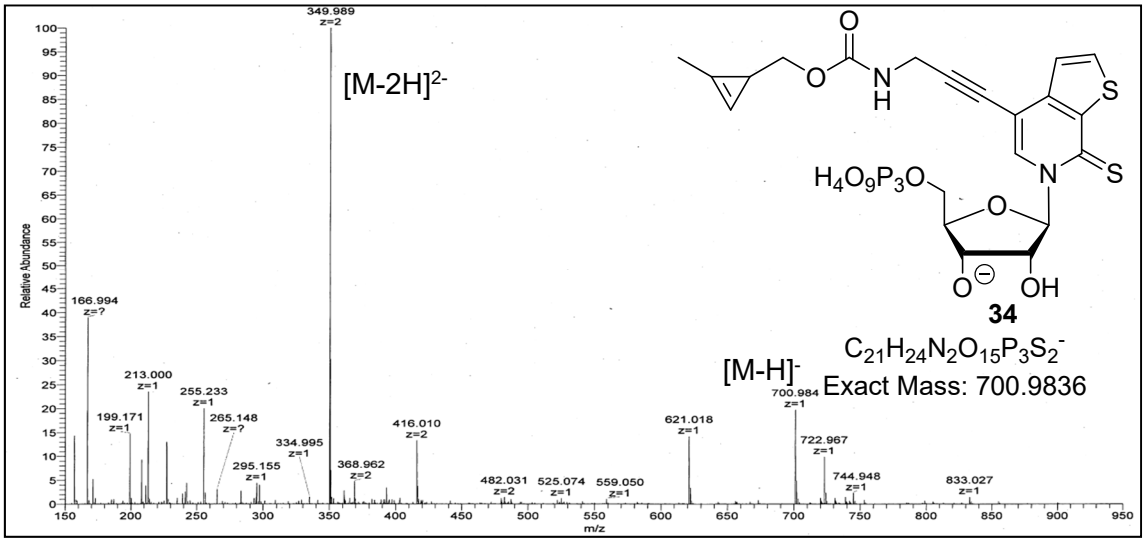


Figure 276. ESI$^-$ MS spectrum of compound **34**.

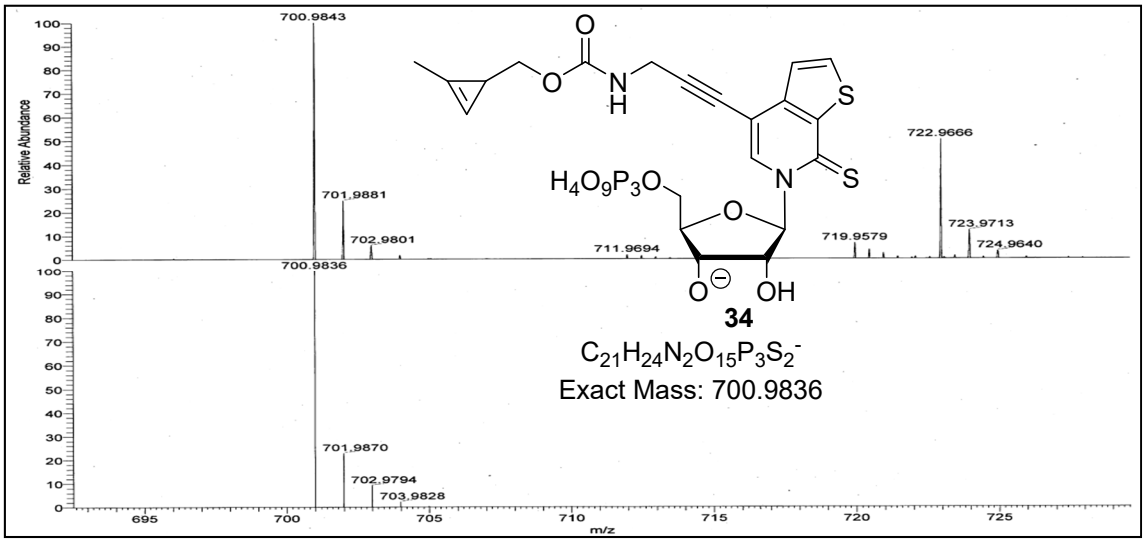


Figure 277. ESI⁻ HR MS spectrum of compound **34**.

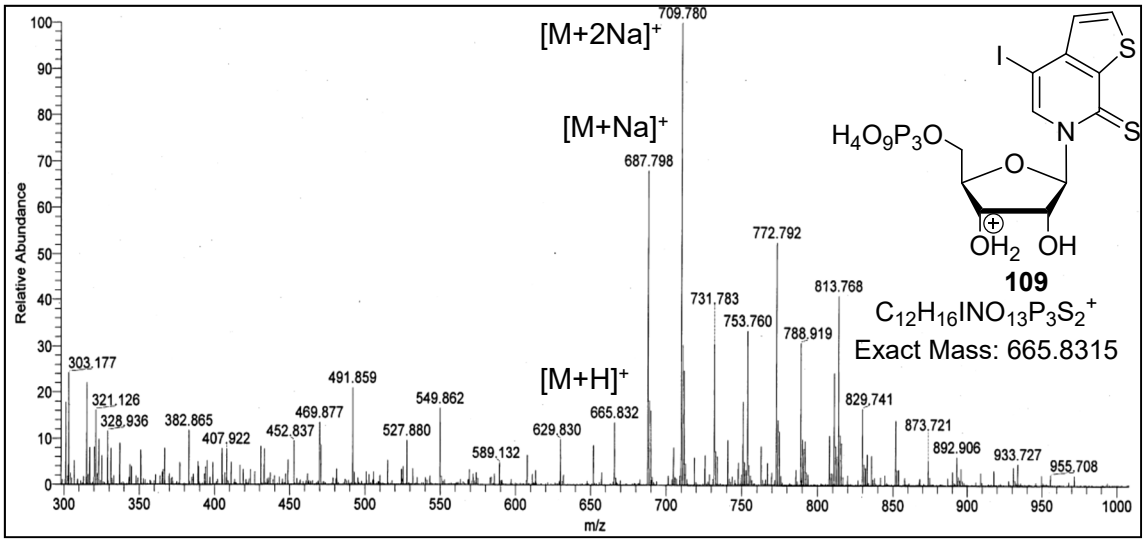


Figure 278. ESI⁺ MS spectrum of compound **109**.

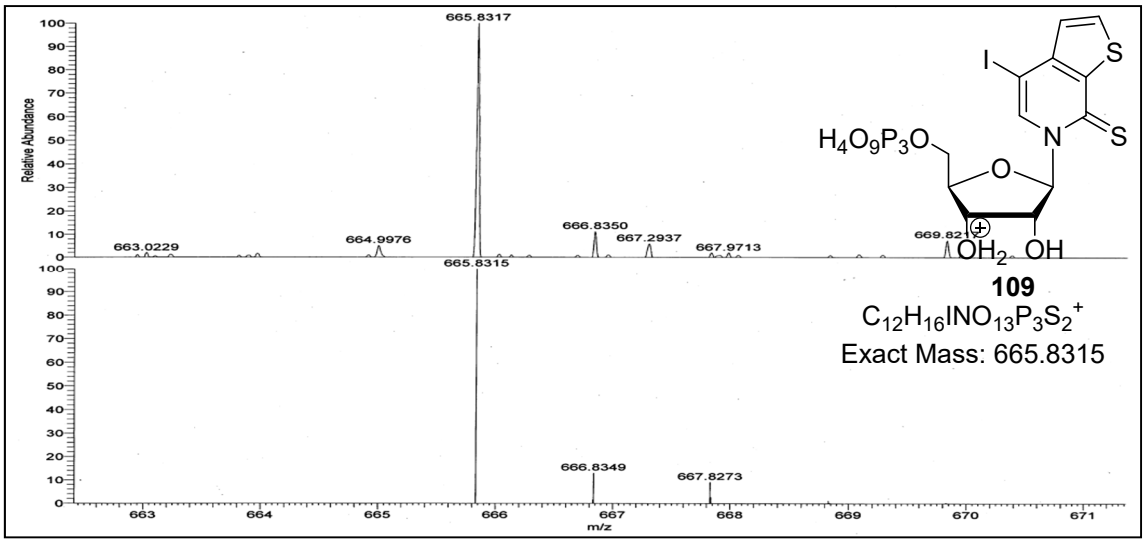


Figure 279. ESI⁺ HR MS spectrum of compound **109**.

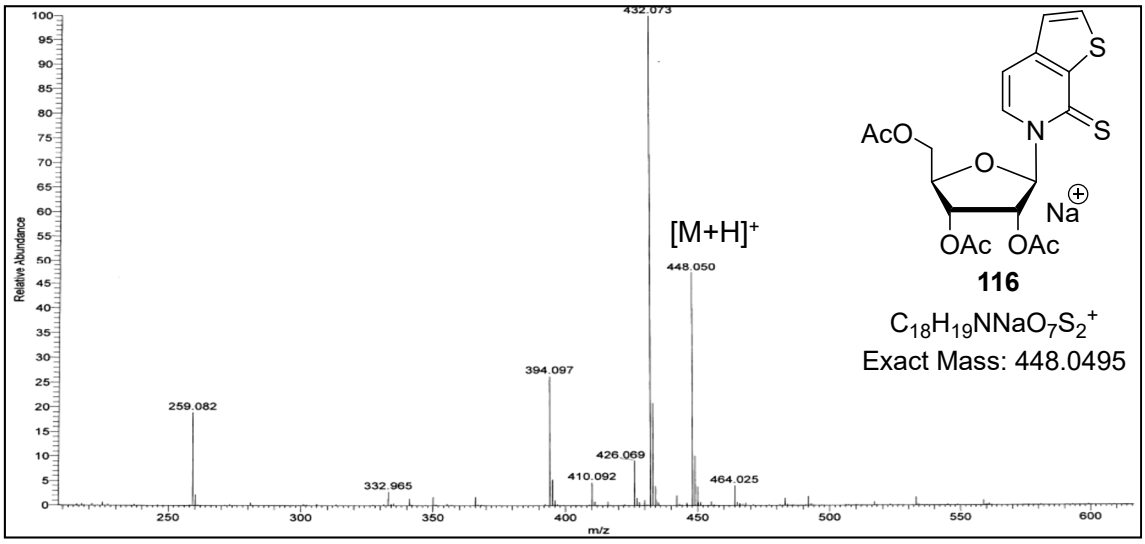


Figure 280. ESI⁺ MS spectrum of compound **116**.

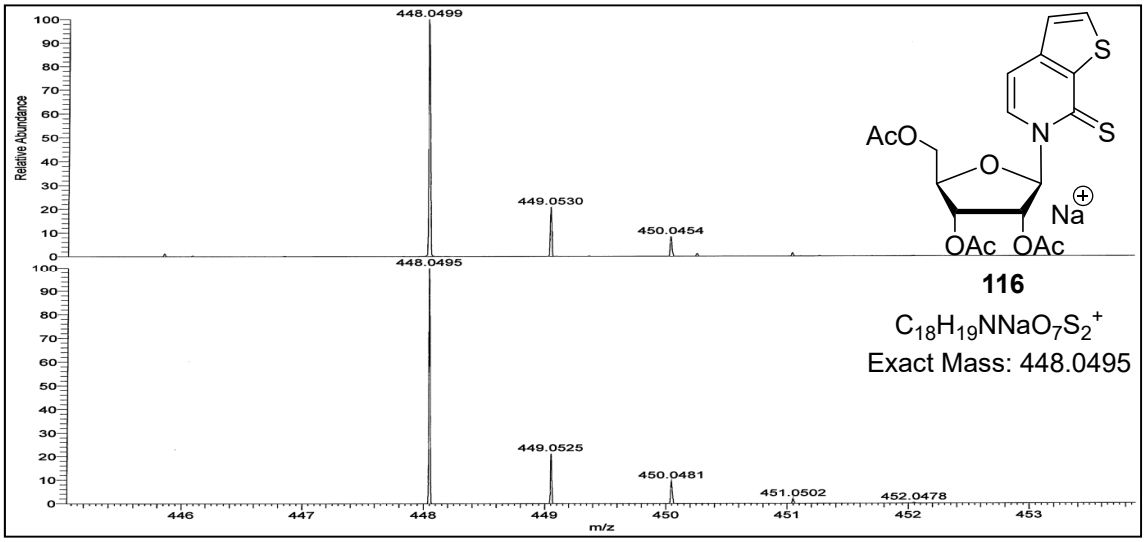


Figure 281. ESI⁺ HR MS spectrum of compound **116**.

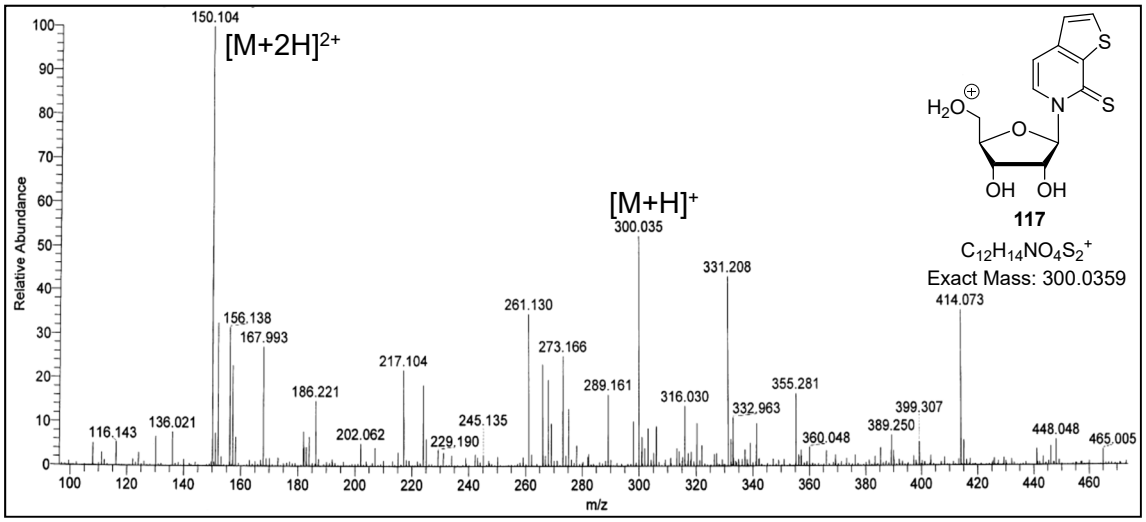


Figure 282. ESI+ MS spectrum of compound **117**.

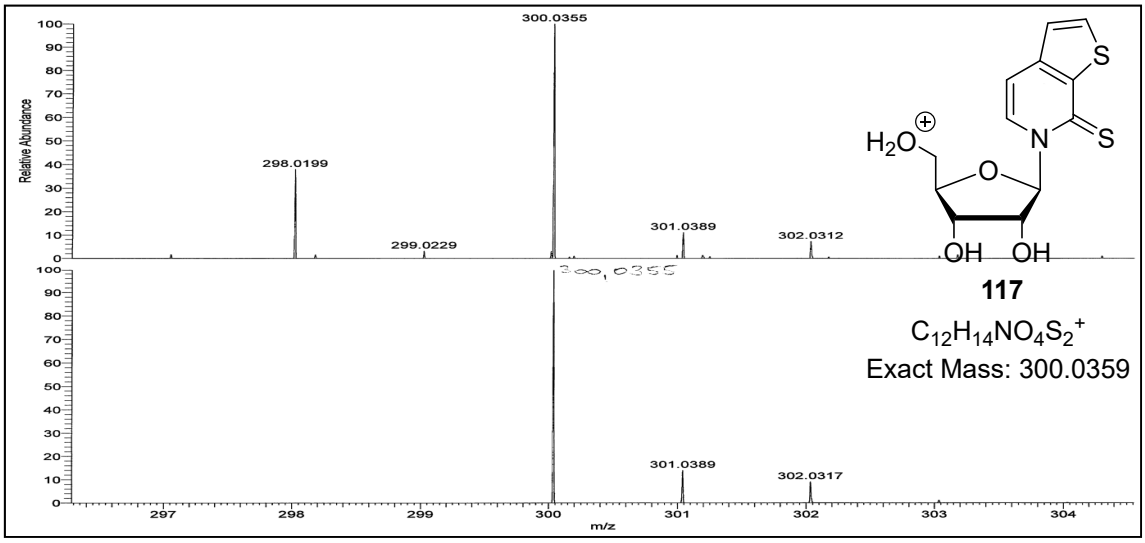


Figure 283. ESI+ HR MS spectrum of compound **117**.

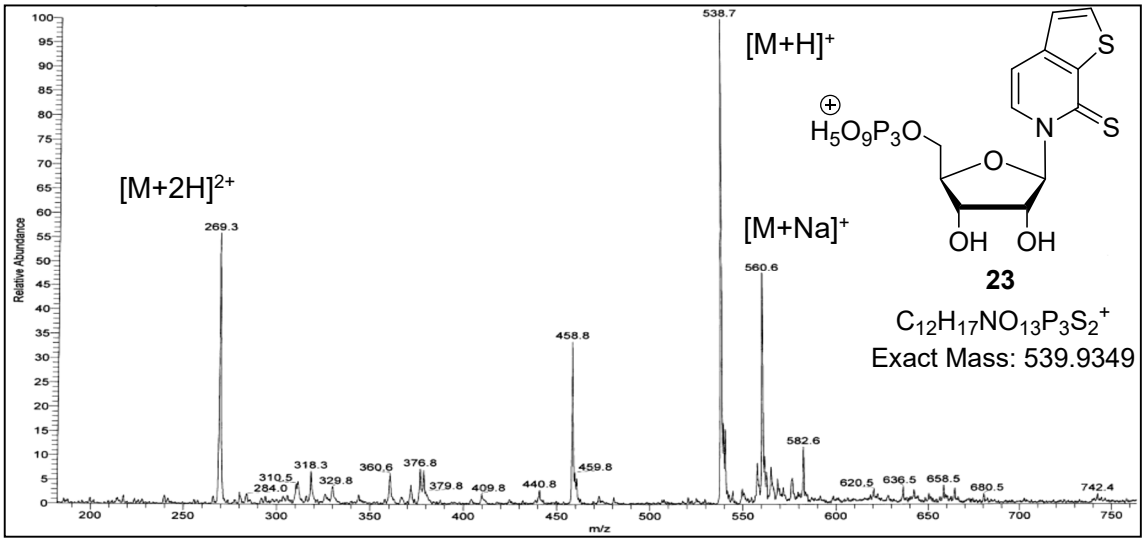


Figure 284. ESI+ MS spectrum of compound **23**.

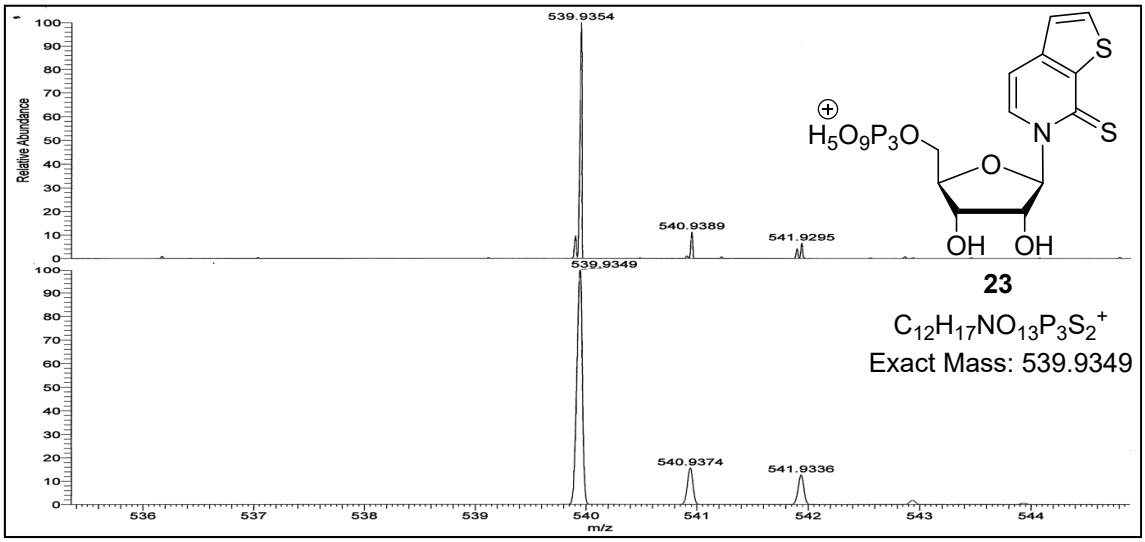


Figure 285. ESI⁺ HR MS spectrum of compound **23**.

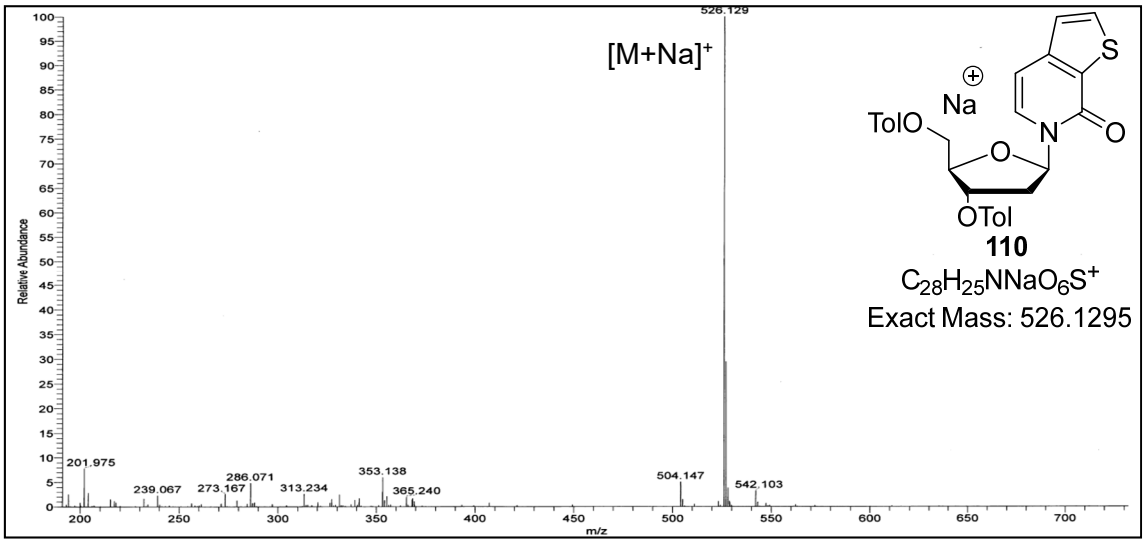


Figure 286. ESI⁺ MS spectrum of compound **110**.

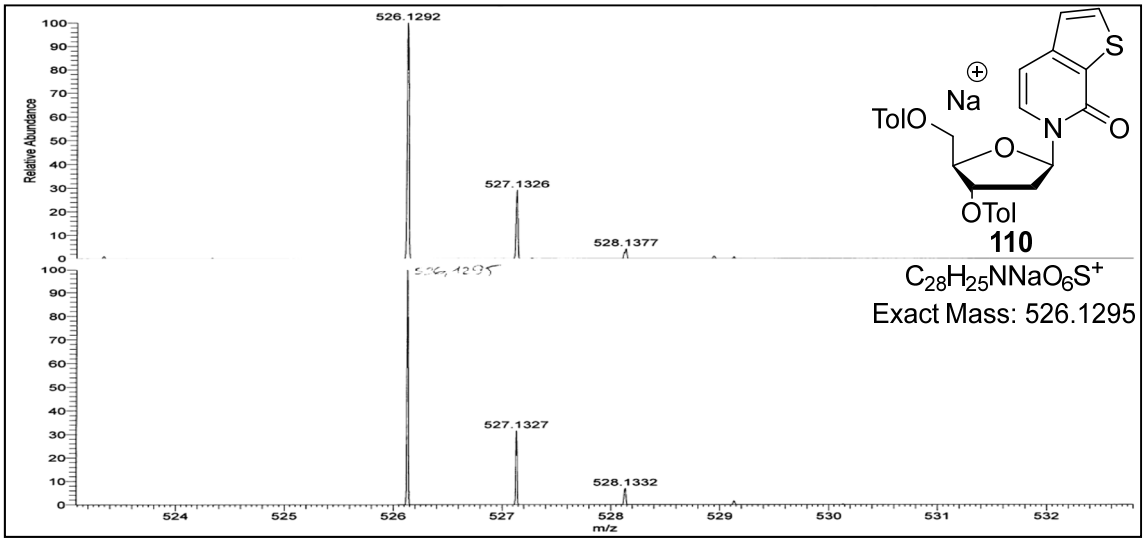
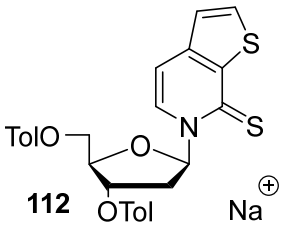


Figure 287. ESI⁺ HR MS spectrum of compound **110**.



$C_{28}H_{25}NNaO_5S_2^+$
Exact Mass: 542.1066

$[M+Na]^+$

Figure 288. ESI$^+$ MS spectrum of compound **112**.

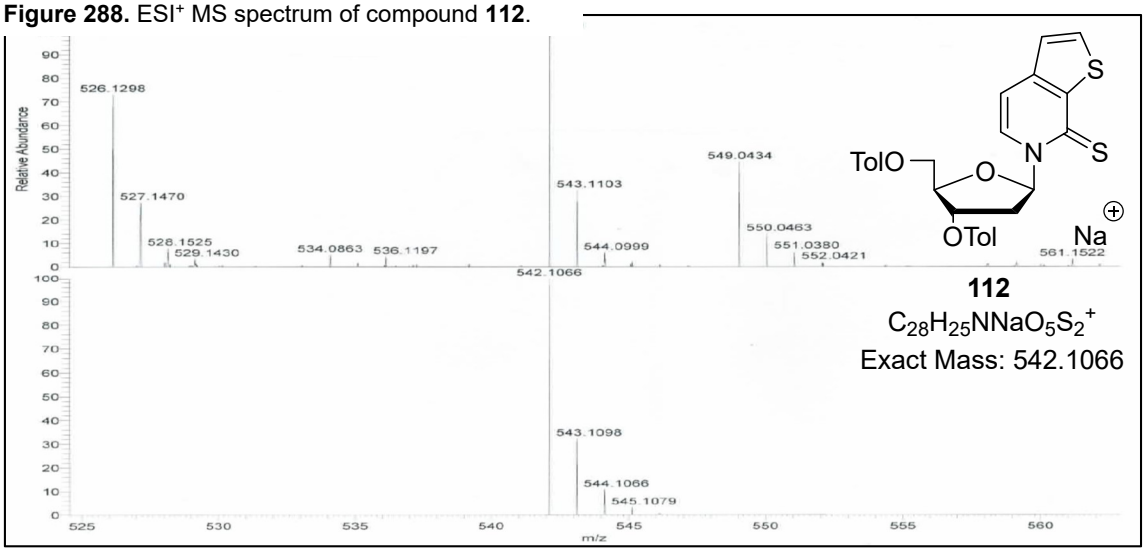


Figure 289. ESI$^+$ HR MS spectrum of compound **112**.

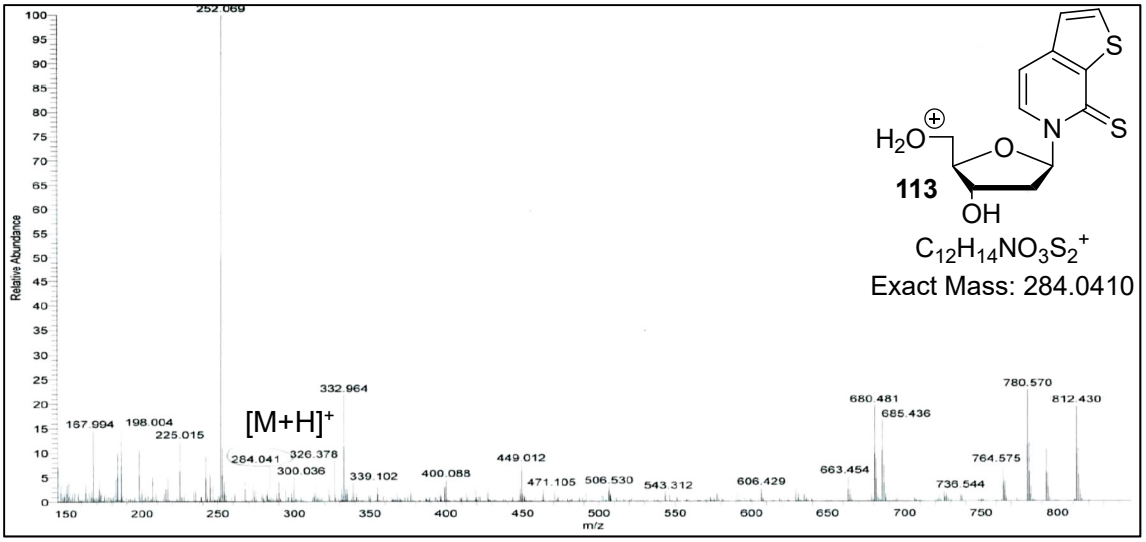
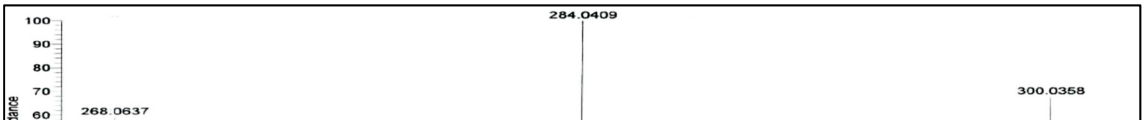
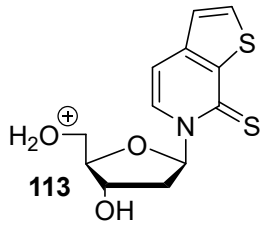


Figure 290. ESI$^+$ MS spectrum of compound **113**.





$C_{12}H_{14}NO_3S_2^+$
Exact Mass: 284.0410

Figure 291. ESI^+ HR MS spectrum of compound **113**.

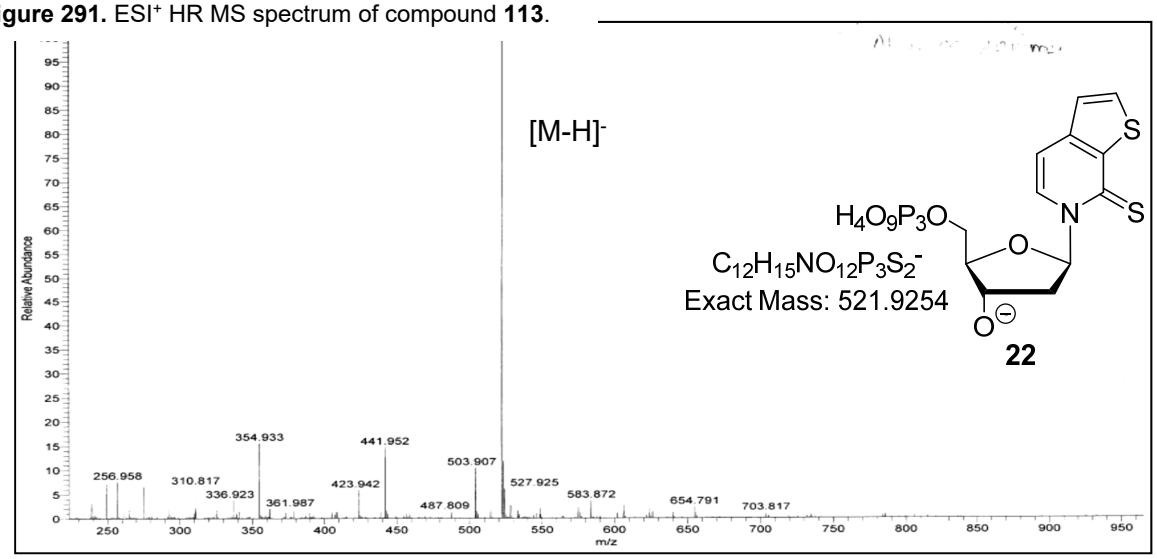


Figure 292. ESI^- MS spectrum of compound **22**.

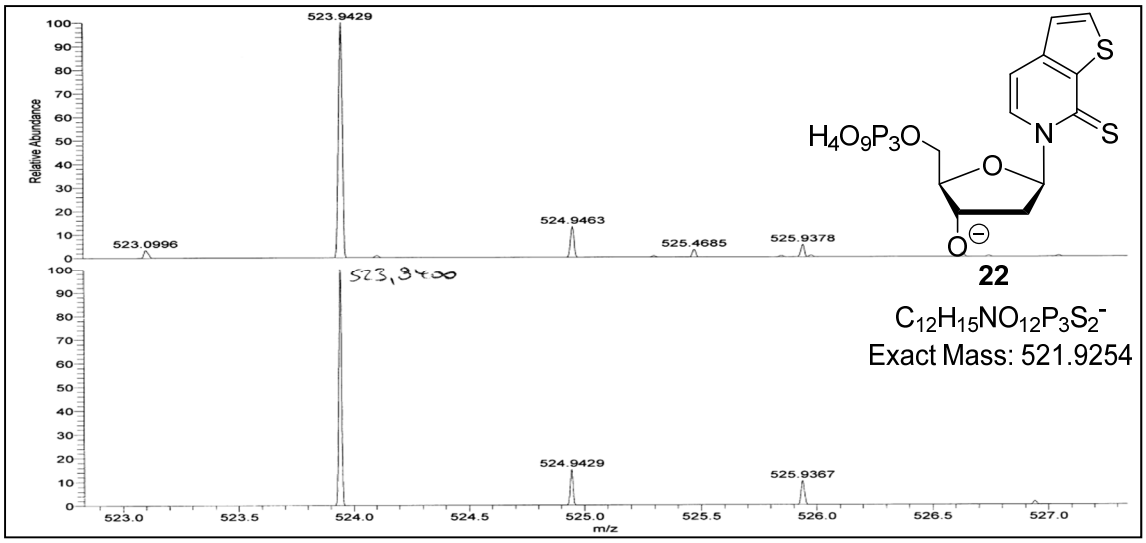


Figure 293. ESI$^-$ HR MS spectrum of compound **22**.

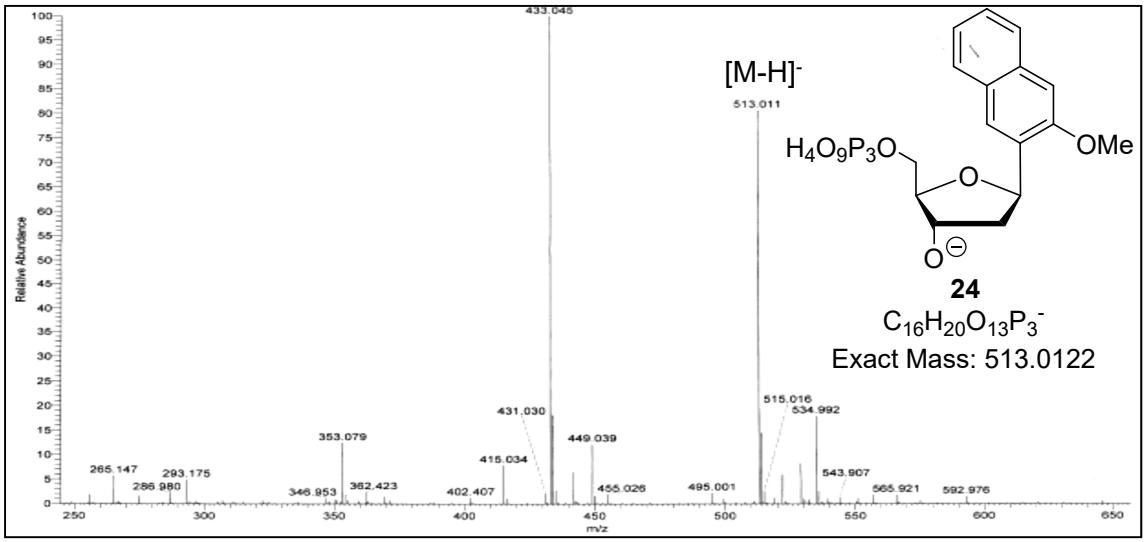


Figure 294. ESI$^-$ MS spectrum of compound **24**.

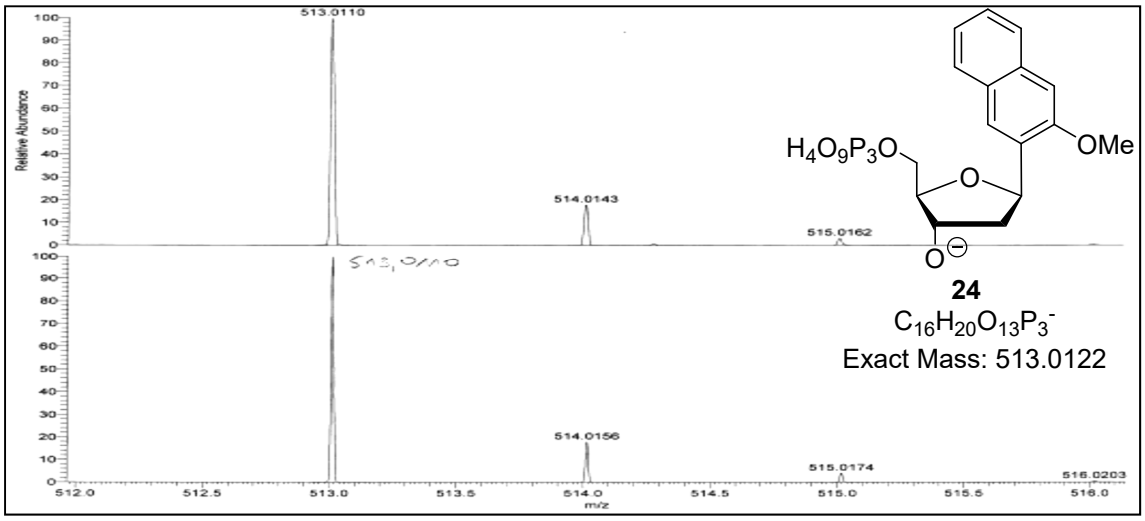


Figure 295. ESI$^-$ HR MS spectrum of compound **24**.

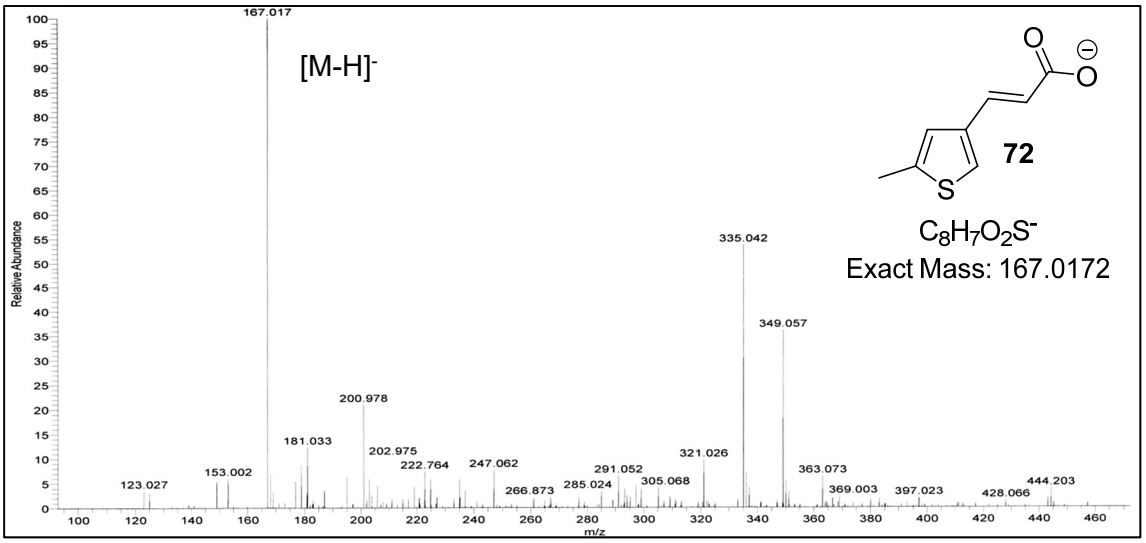


Figure 296. ESI$^-$ MS spectrum of compound **72**.

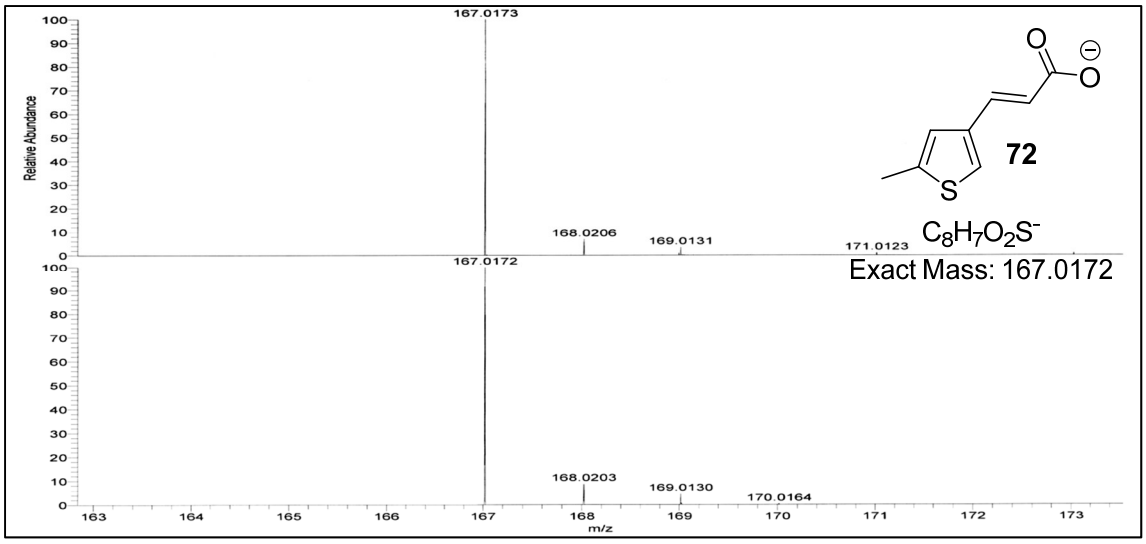


Figure 297. ESI$^-$ HR MS spectrum of compound **72**.

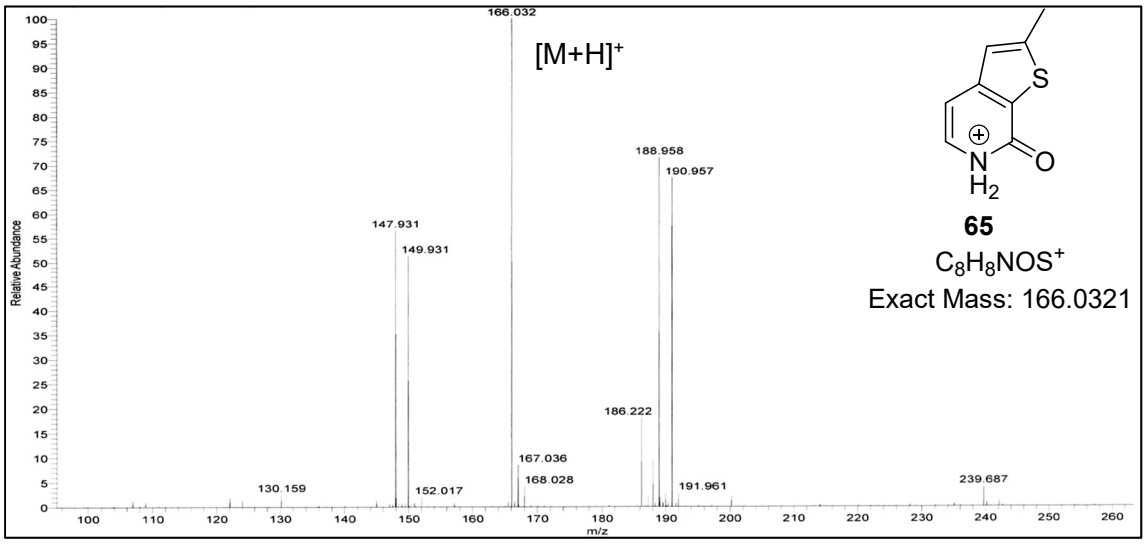


Figure 298. ESI$^+$ MS spectrum of compound **65**.

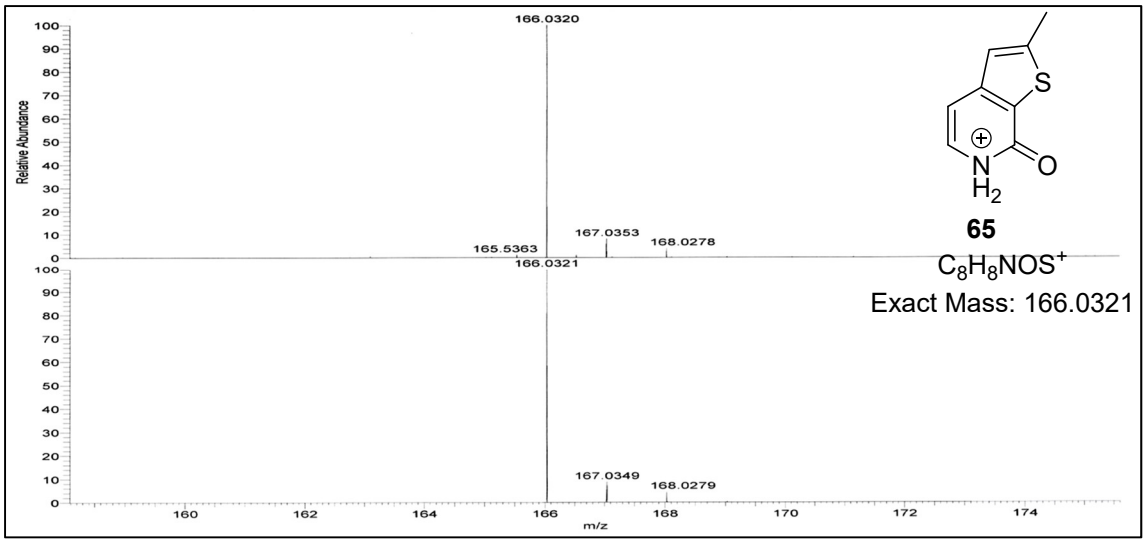


Figure 299. ESI⁺ HR MS spectrum of compound **65**.

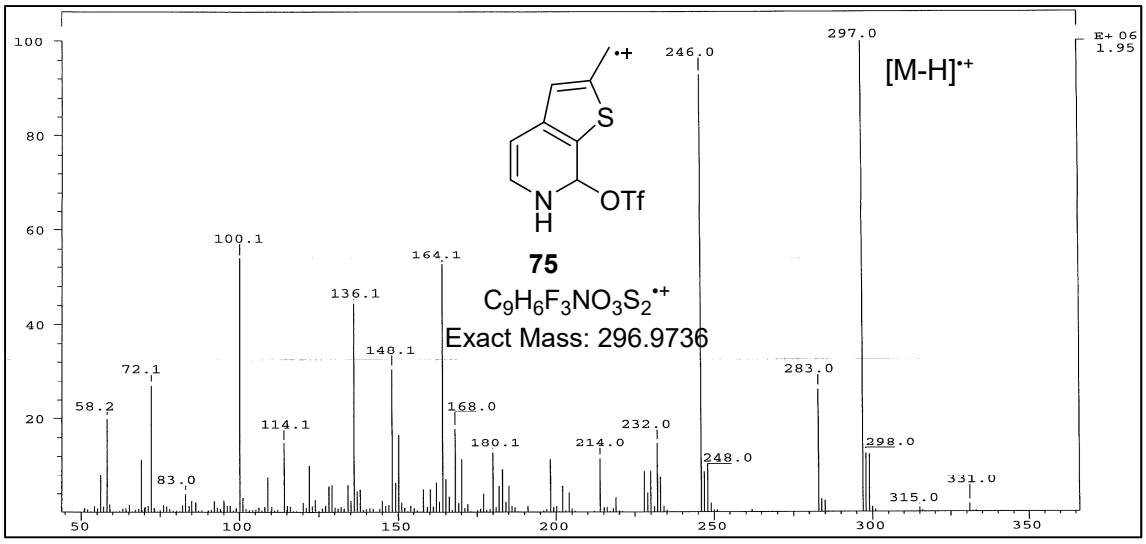


Figure 300. EI MS spectrum of compound **75**.

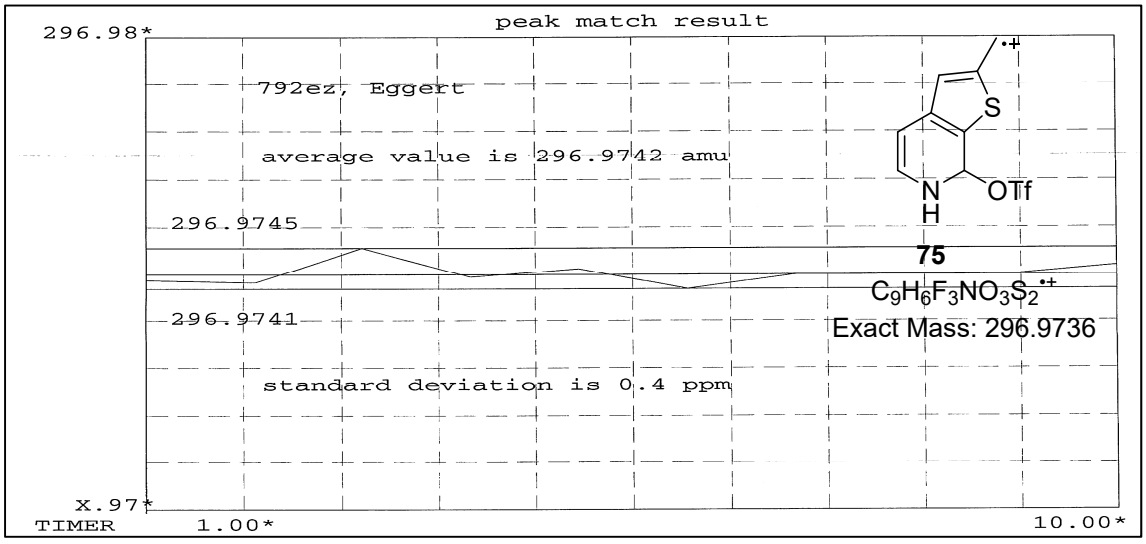


Figure 301. EI HR MS spectrum of compound **75**.

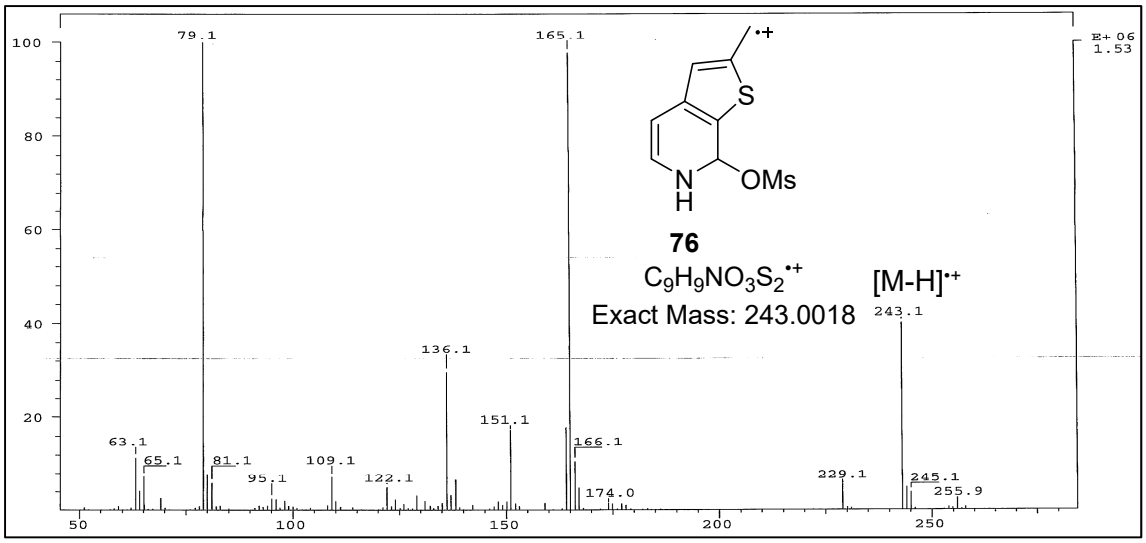


Figure 302. EI MS spectrum of compound **76**.

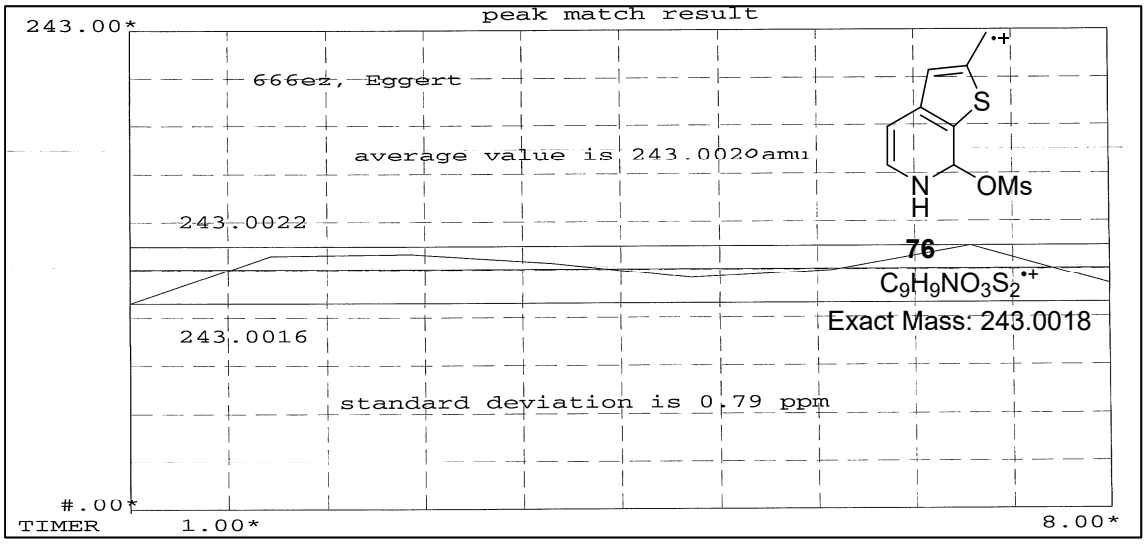


Figure 303. EI HR MS spectrum of compound **76**.

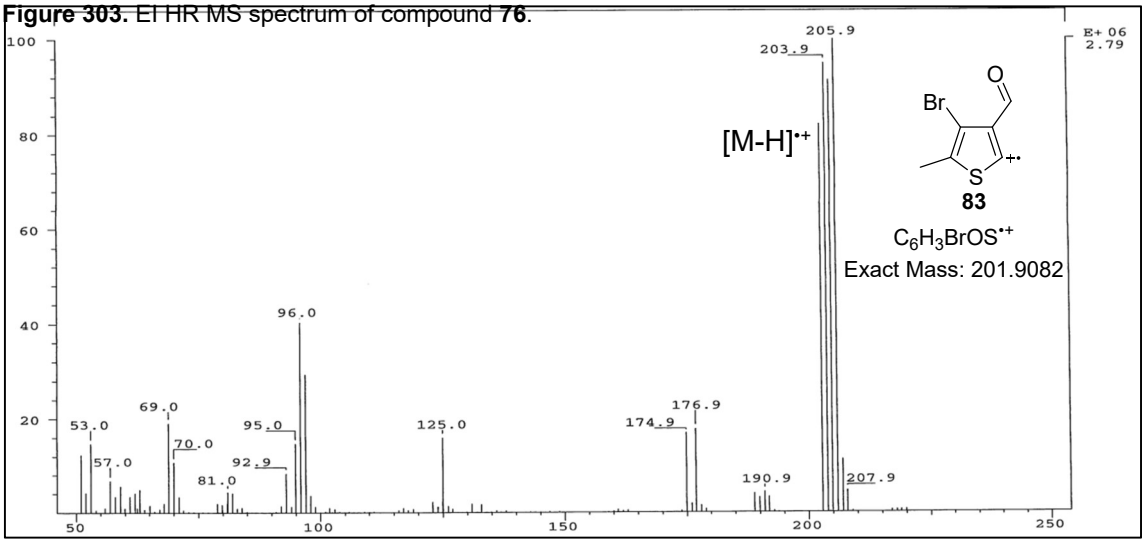


Figure 304. EI MS spectrum of compound **83**.

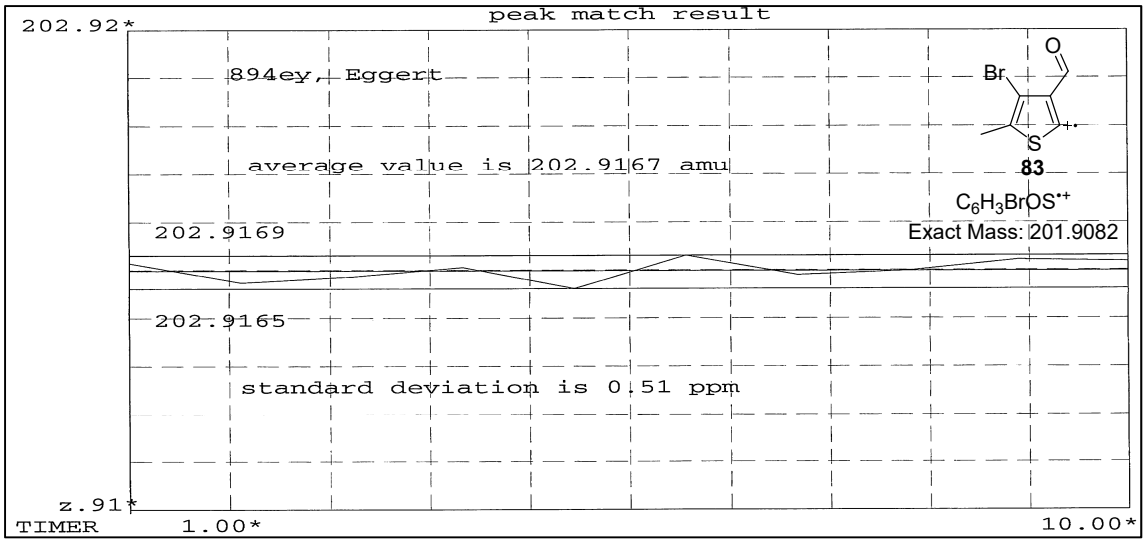


Figure 305. EI HR MS spectrum of compound **83**.

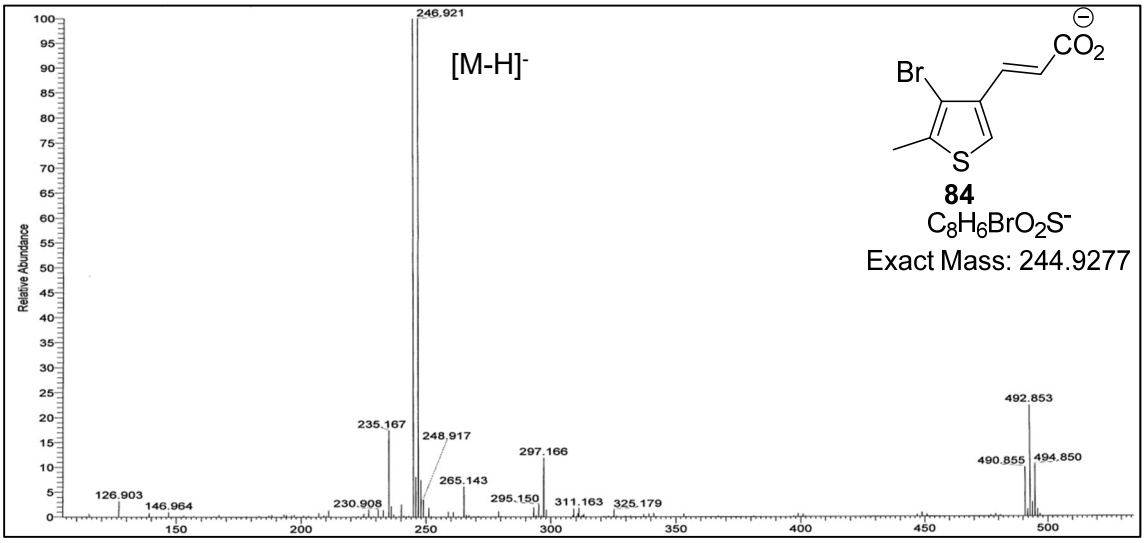


Figure 306. ESI- MS spectrum of compound **84**.

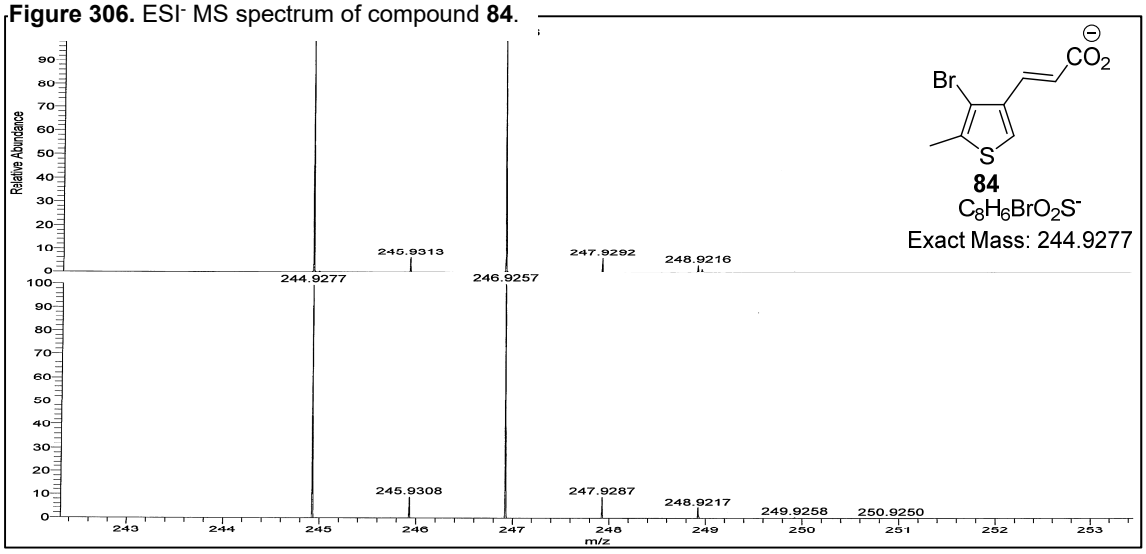
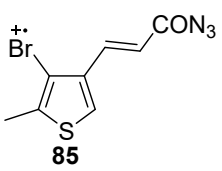


Figure 307. ESI- HR MS spectrum of compound **84**.



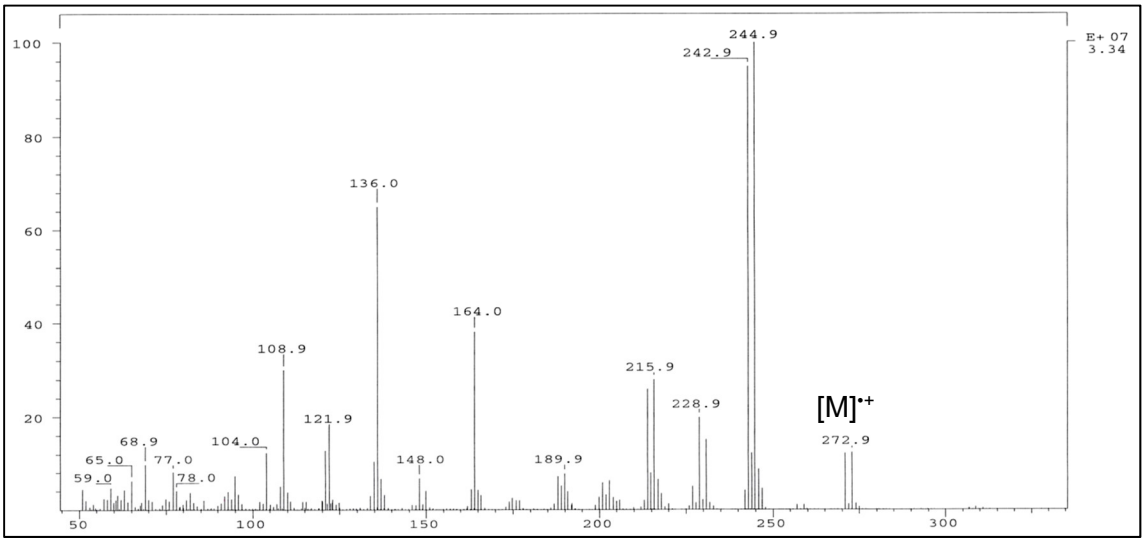


Figure 308. EI MS spectrum of compound **85**.

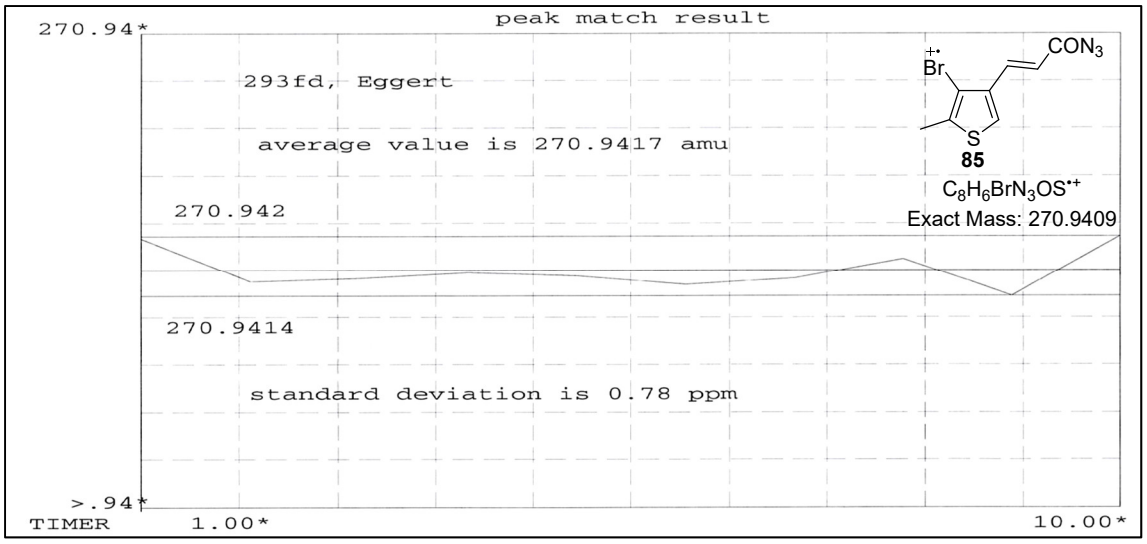


Figure 309. EI HR MS spectrum of compound **85**.

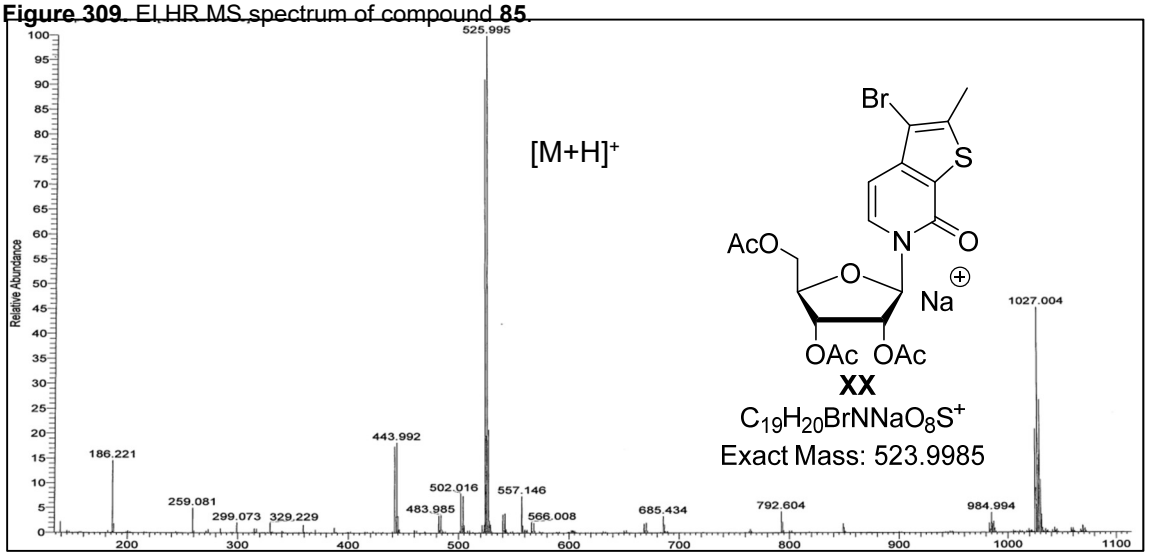


Figure 310. ESI+ MS spectrum of compound **104**.

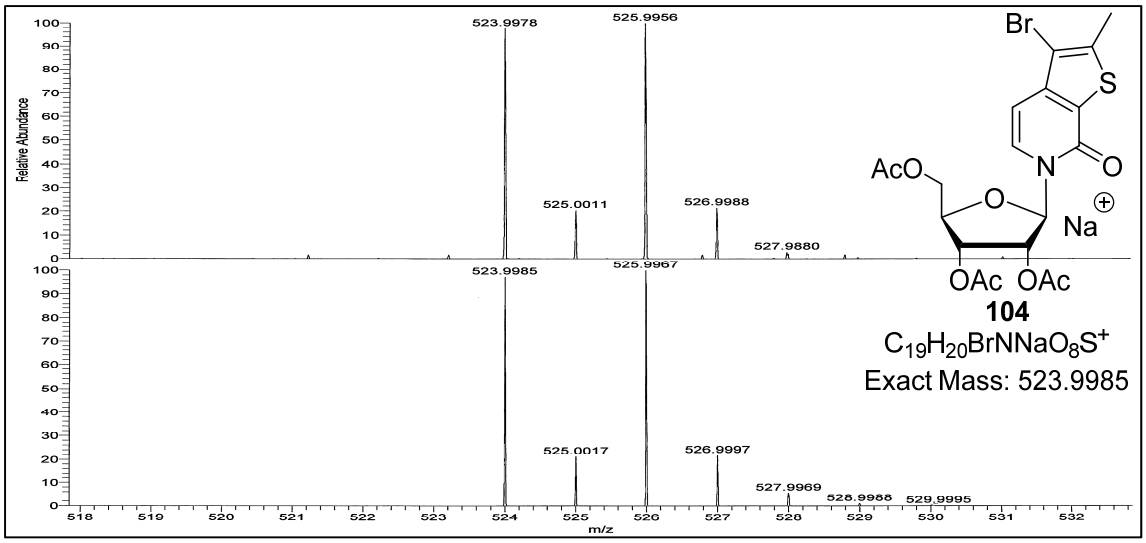


Figure 311. ESI⁺ HR MS spectrum of compound **104**.

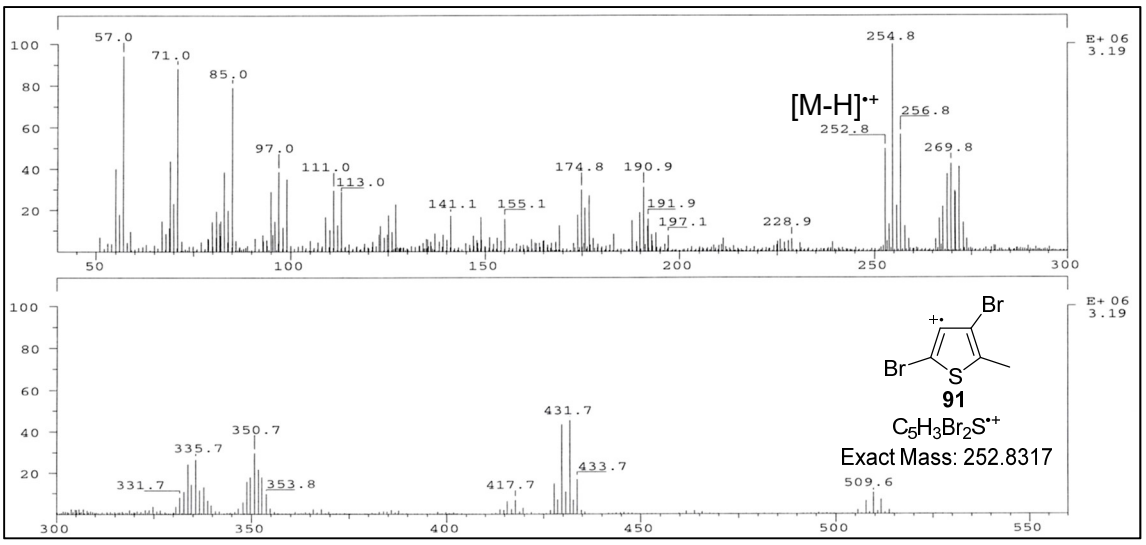


Figure 312. EI MS spectrum of compound **91**.

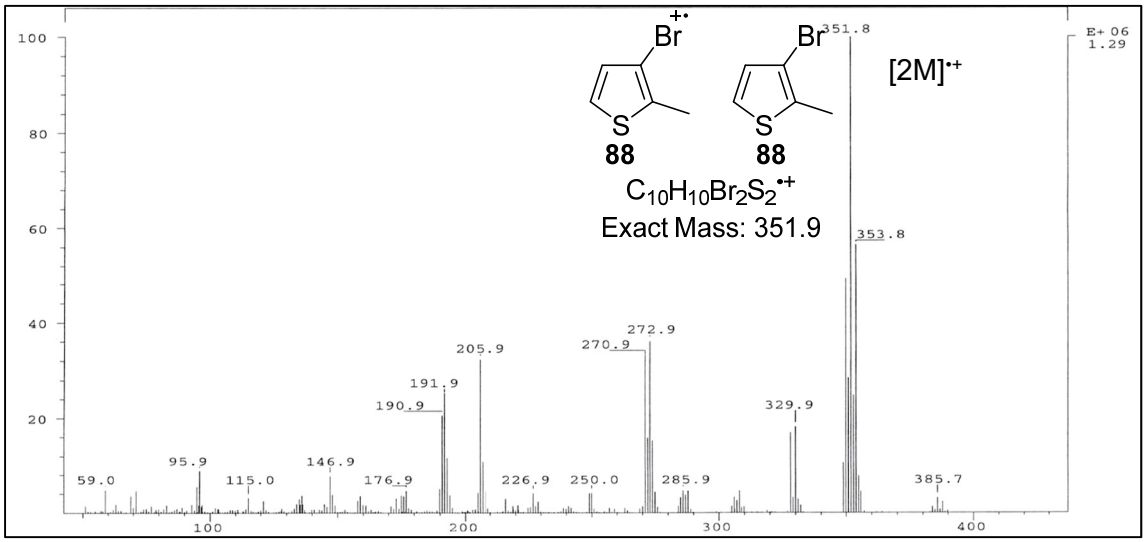


Figure 313. EI MS spectrum of compound **88**.

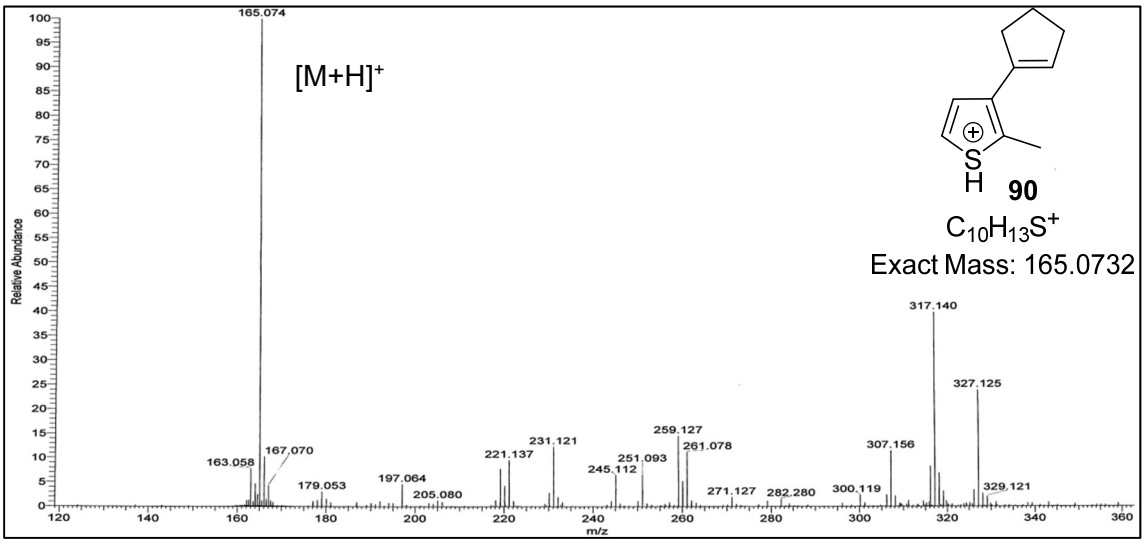


Figure 314. ESI+ MS spectrum of compound **90**.

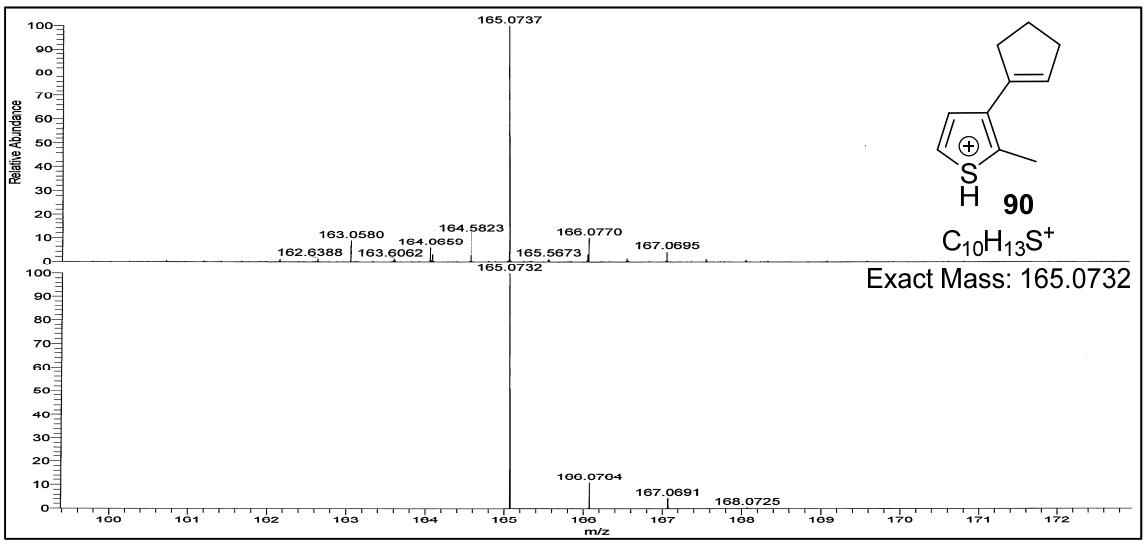


Figure 315. ESI+ HR MS spectrum of compound **90**.

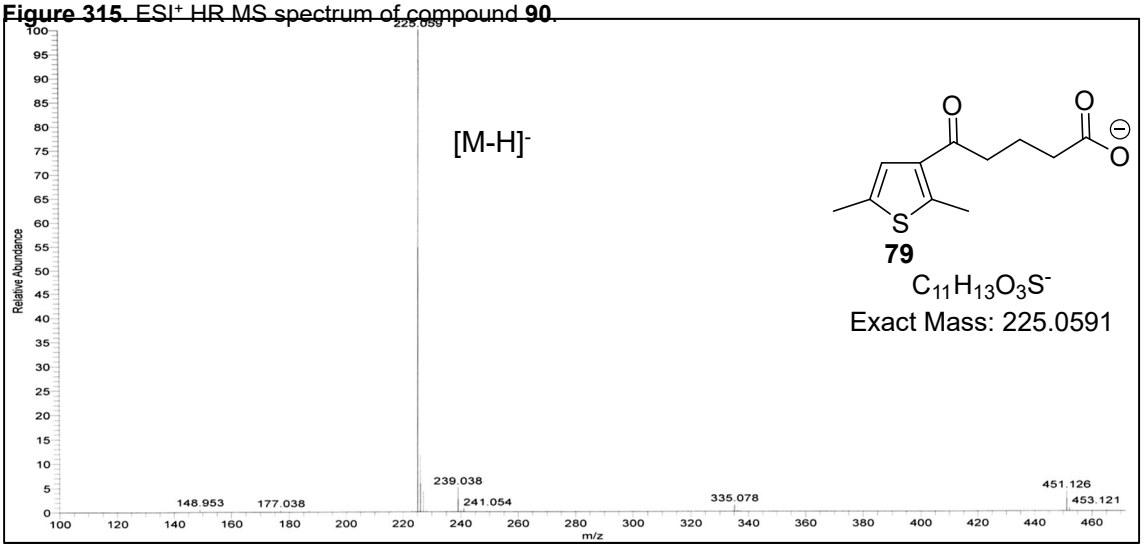


Figure 316. ESI- MS spectrum of compound **79**.

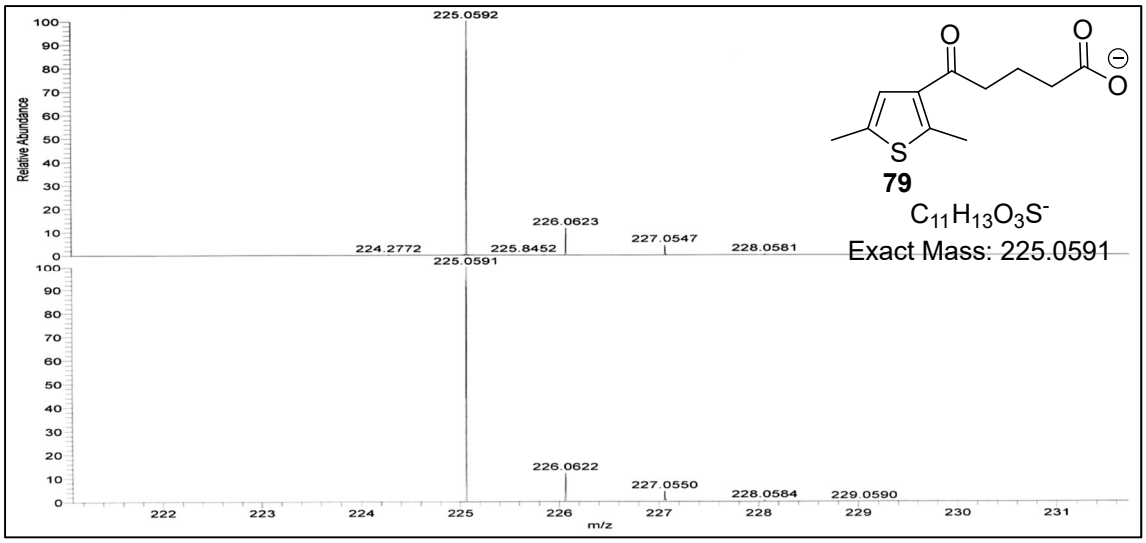


Figure 317. ESI$^-$ HR MS spectrum of compound **79**.

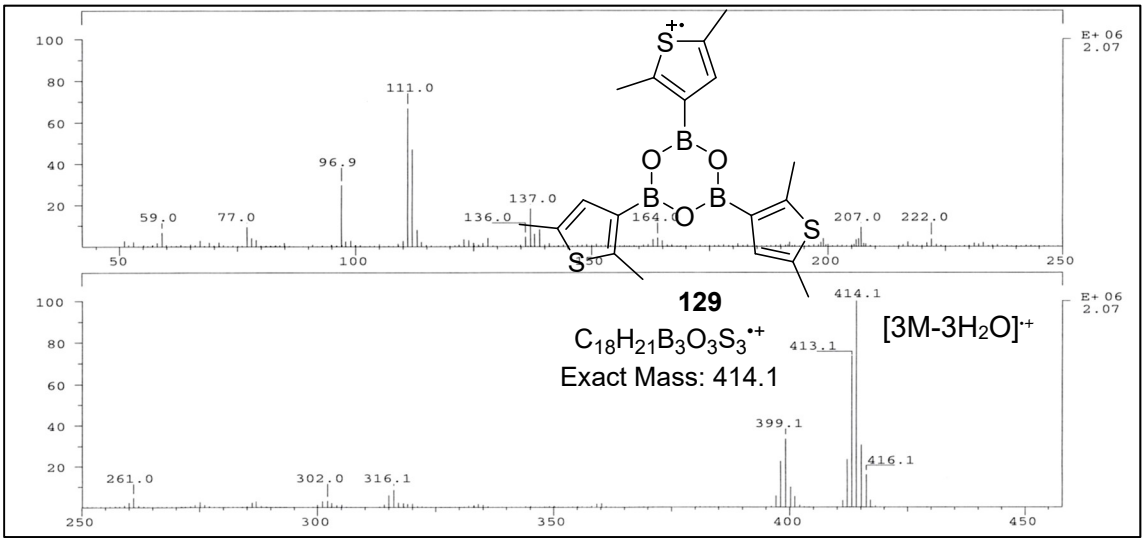


Figure 318. EI MS spectrum of compound **94** or its trimer **129**.

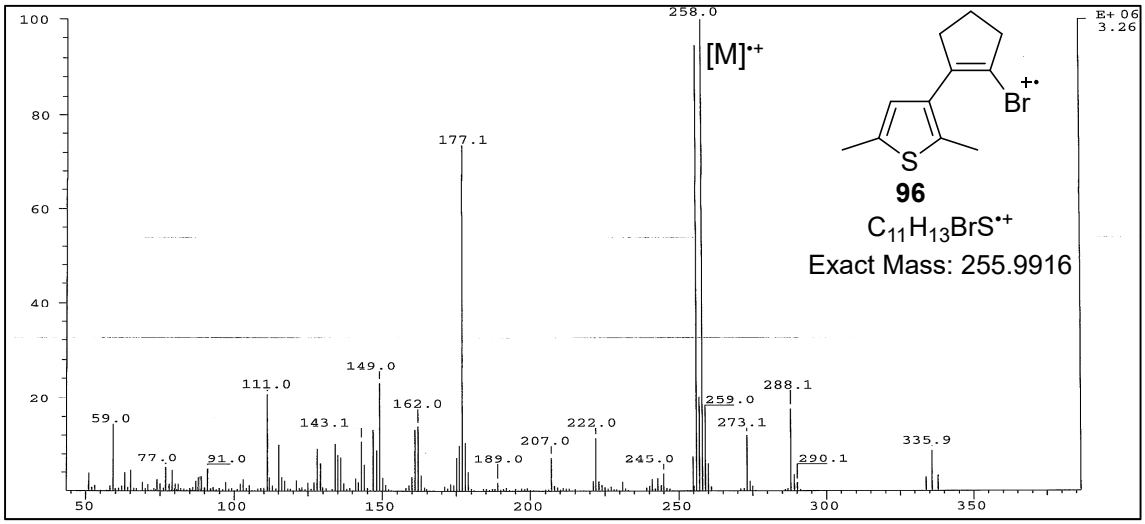


Figure 319. EI MS spectrum of compound **96**.

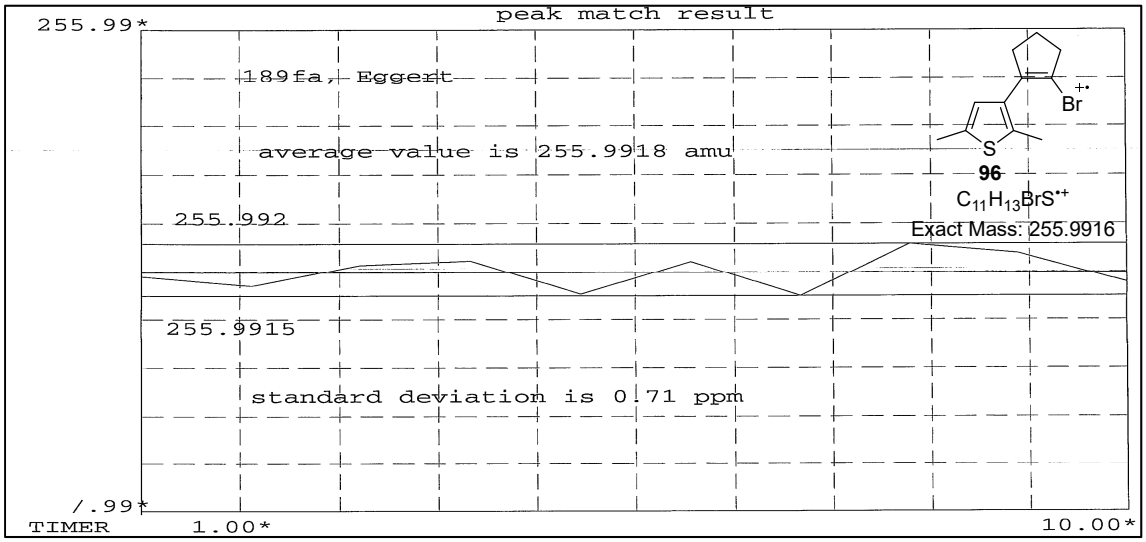


Figure 320. EI HR MS spectrum of compound **96**.

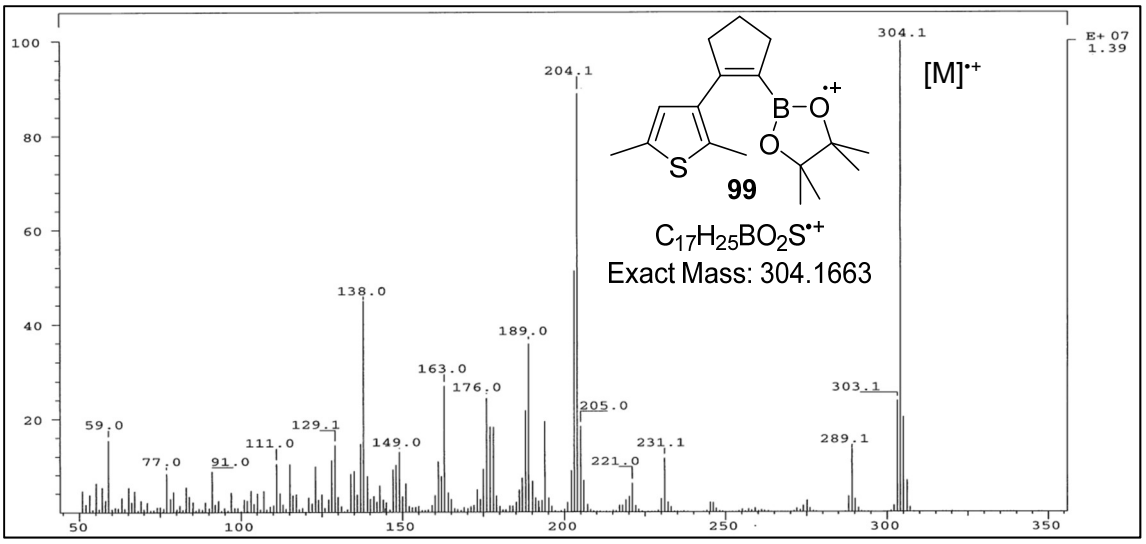


Figure 321. EI MS spectrum of compound **99**.

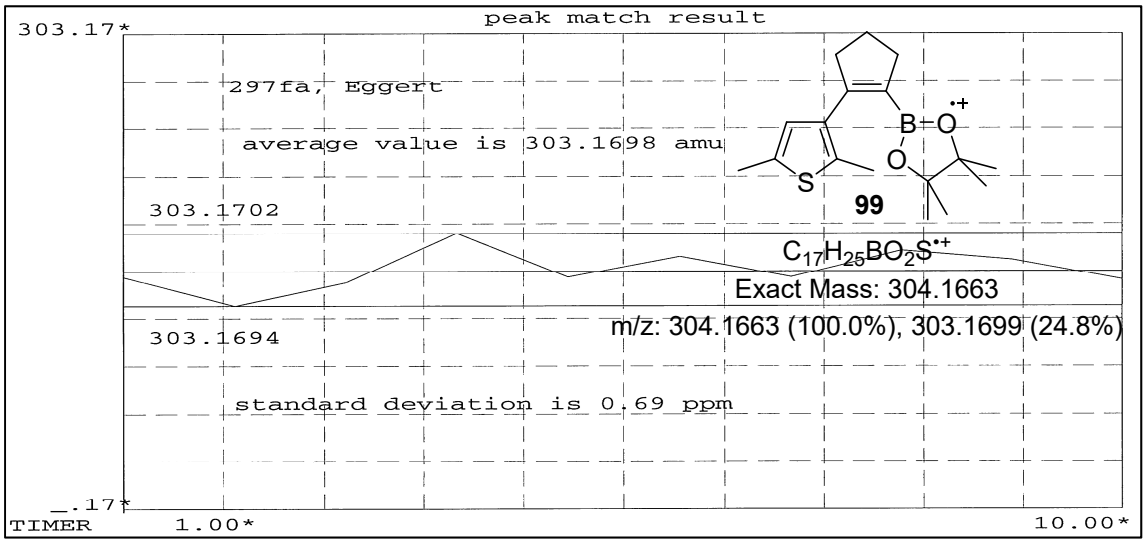


Figure 322. EI HR MS spectrum of compound **99**.

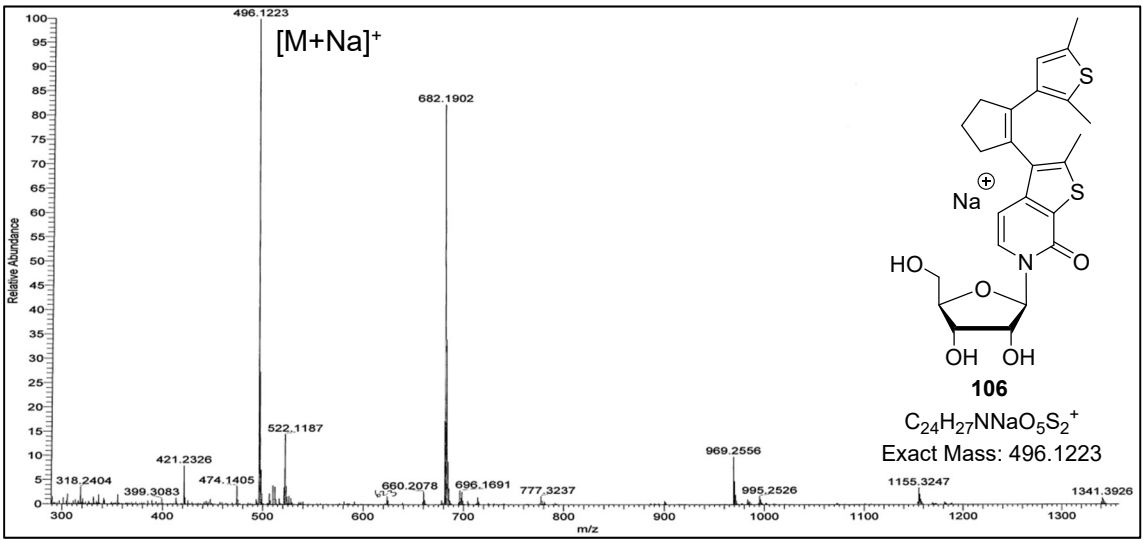


Figure 323. ESI+ MS spectrum of compound **106**.

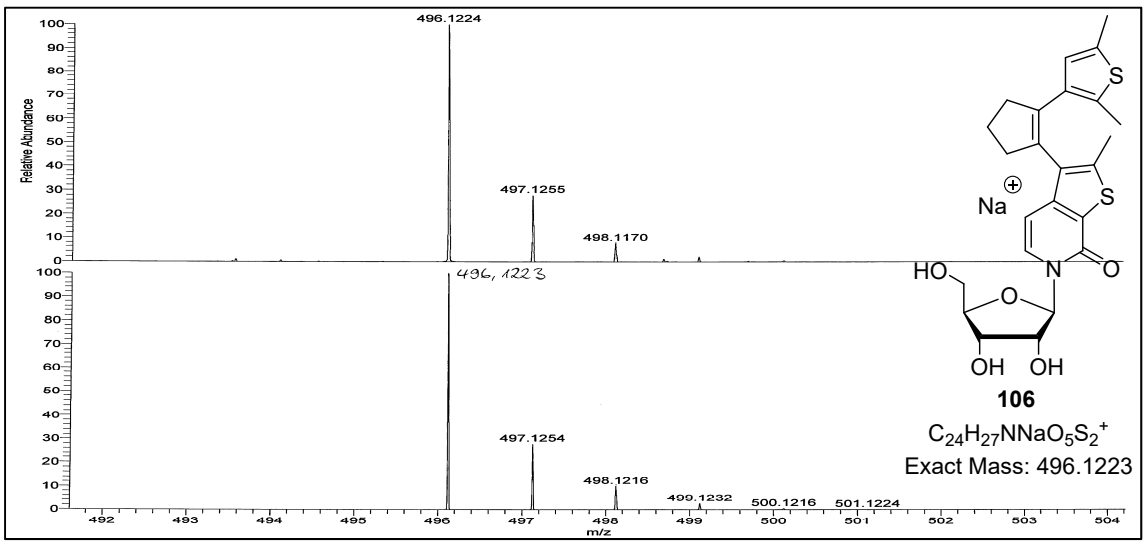


Figure 324. ESI+ HR MS spectrum of compound **106**.

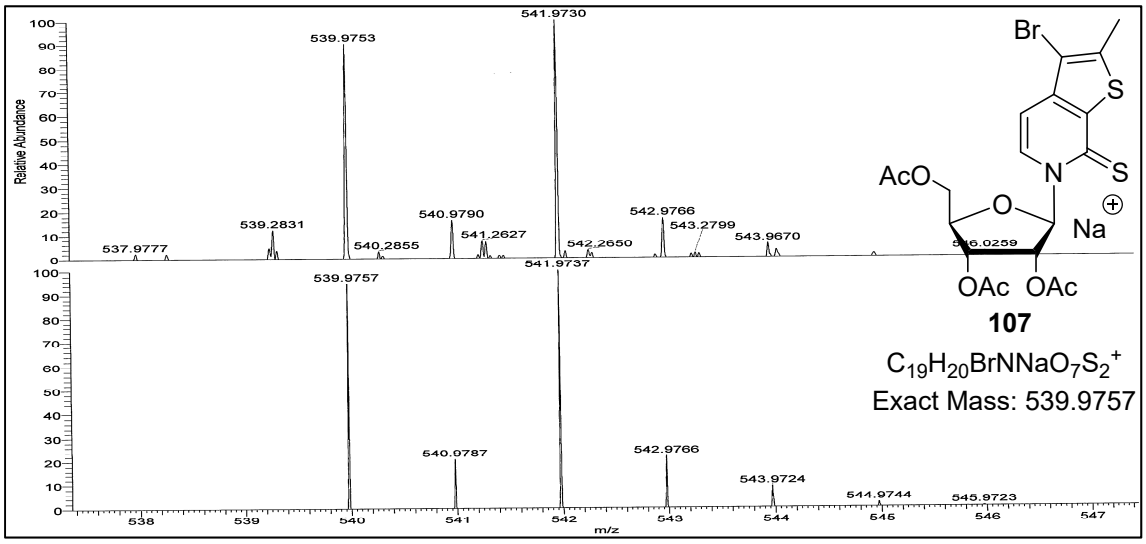


Figure 325. ESI+ HR MS spectrum of compound **107**.

7.4 Gel electrophoresis scans

This section includes all full-lane images of DPAGE and agarose gels, *i.e.* UV scans and fluorescence scans depicted in the main text.

DPAGE

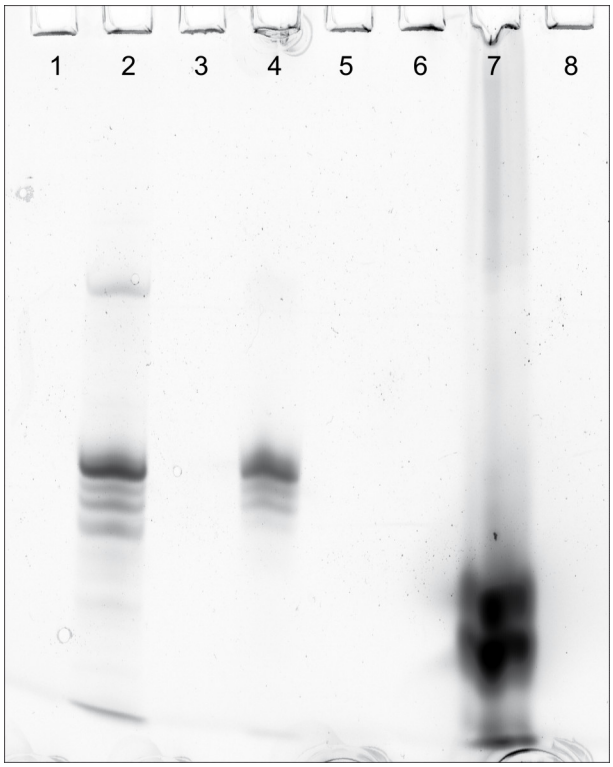


Figure 326. Full lanes of the 20 % DPAGE fluorescence scan depicted in Figure 53; Lane 1: crude **RNA^{CP}**; Lane 2: crude **RNA^{CP}** incubated with **28**; Lane 3: purified **RNA^{CP}**; Lane 4: purified **RNA^{CP}** incubated with **28**; Lane 5: crude **RNA^{n1}**; Lane 6: crude **RNA^{n1}** incubated with **28**; Lane 7: tetrazine-Oregon Green conjugate **28**; Lane 8: template **DNA^{n1}**.

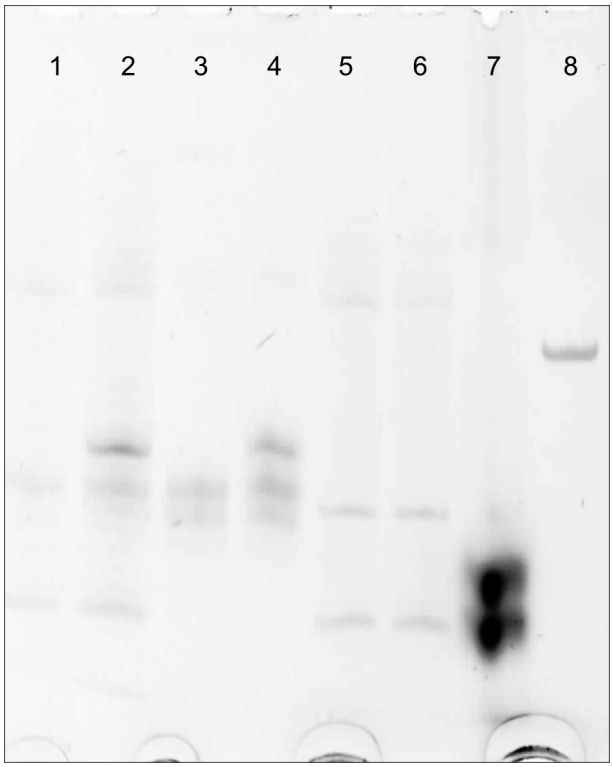


Figure 327. Full lanes of the 20 % DPAGE UV scan (*SYBR Safe*-stained) depicted in Figure 53; Lane 1: DNase-digested **RNA^{CP}**; Lane 2: DNase-digested **RNA^{CP}** incubated with **28**; Lane 3: purified **RNA^{CP}**; Lane 4: purified **RNA^{CP}** incubated with **28**; Lane 5: DNase-digested **RNA^{n1}**; Lane 6: DNase-digested **RNA^{n1}** incubated with **28**; Lane 7: tetrazine-Oregon Green conjugate **28**; Lane 8: template **DNA^{n1}**.

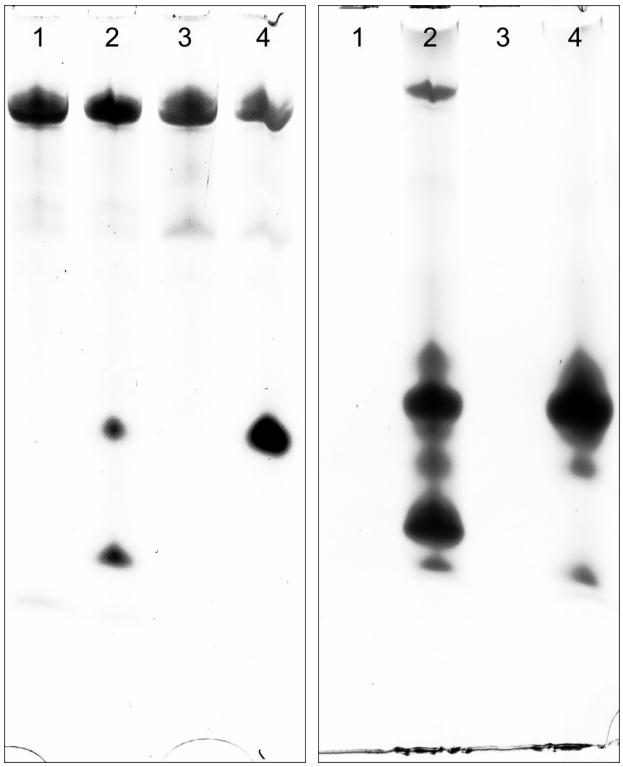


Figure 328. Full lanes of the 20 % DPAGE fluorescence scan (left panel) and UV scan (*SYBR Safe*-stained, right panel) depicted in Figure 55; Lane 1: DNase-digested **tRNA^{CP}**; Lane 2: DNase-digested **tRNA^{CP}** incubated with **28**; Lane 3: DNase-digested **tRNA^{n}**; Lane 4: DNase-digested **tRNA^{n}** incubated with **28**.

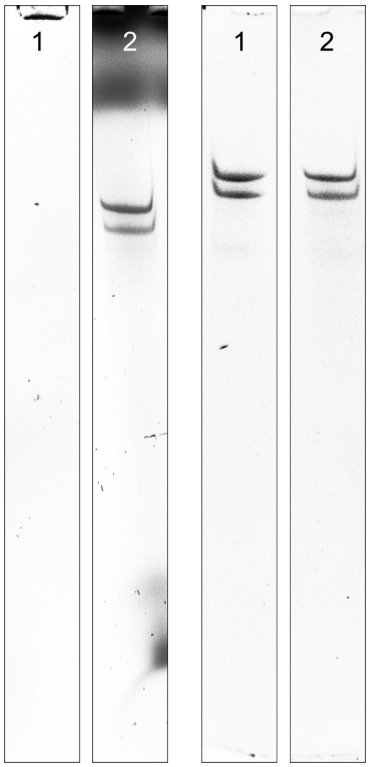


Figure 329. Full lanes of the 20 % DPAGE fluorescence scans (left two panels) and UV scan (*SYBR Safe*-stained, right two panels) depicted in Figure 60; Lane 1: **glmS_RNA^{CP}**, DNase-digested and gel-filtrated; Lane 2: DNase-digested and gel-filtrated **glmS_RNA^{CP}**, incubated with **57**.

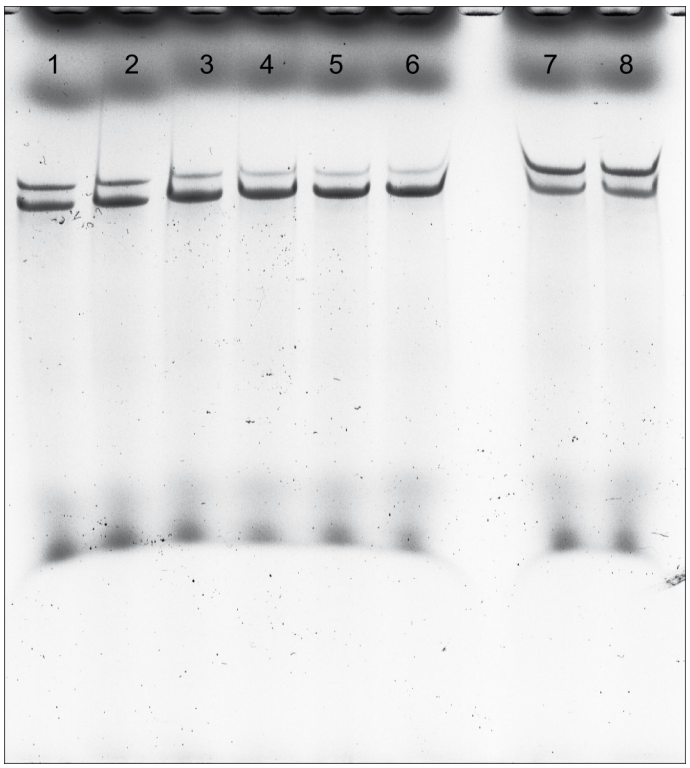


Figure 330. Full lanes of the 20 % DPAGE fluorescence scan depicted in Figure 61; cleavage assay of **glmS_RNA^{CP}-Tet-OG** incubated with **1** after 1 min (Lane 1), 2 min (Lane 2), 5 min (Lane 3), 15 min (Lane 4), 30 min (Lane 5), and 60 min (Lane 6); Lane 7: **glmS_RNA^{CP}-Tet-OG** after 1 min; Lane 8: **glmS_RNA^{CP}-Tet-OG** after 60 min.

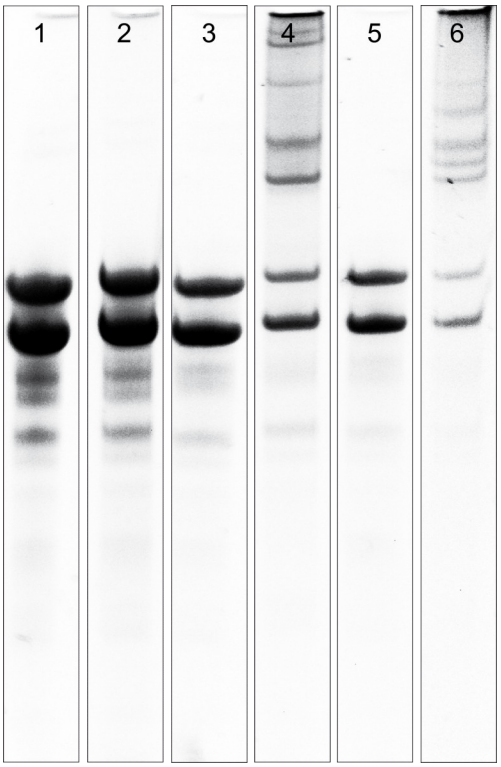


Figure 331. Full lanes of the 8 % DPAGE (*SYBR Safe* stain) shown in Figure **64**; Lane 1: **glmS_RNA^{n}**; Lane 2: **glmS_RNA^{n}** + **58** and streptavidin; Lane 3: **glmS_RNA^{CP}1**; Lane 4: **glmS_RNA^{CP}1** + **58** and streptavidin; Lane 5: : **glmS_RNA^{CP}**; Lane 6: **glmS_RNA^{CP}** + **58** and streptavidin.

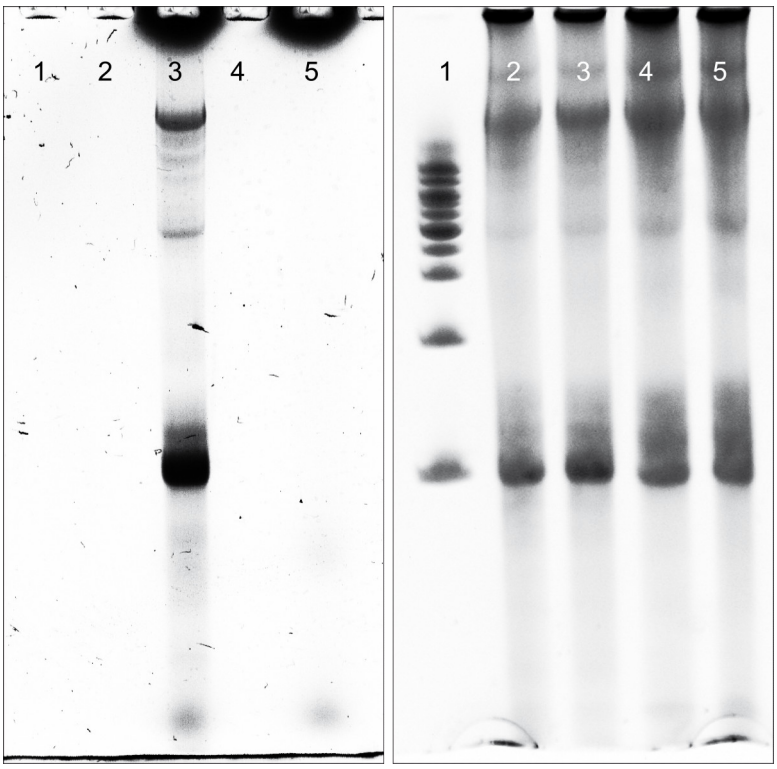


Figure 332. Full lanes of the 20 % DPAGE fluorescence scan (left panel) and UV scan (*SYBR Safe*-stained) depicted in Figure 66; Lane 1: Marker; Lane 2: **Xist_RNA^{CP}**; Lane 3: **Xist_RNA^{CP}+59**; Lane 4: **Xist_RNA^{n}**; Lane 5: : **Xist_RNA^{n}+59**.

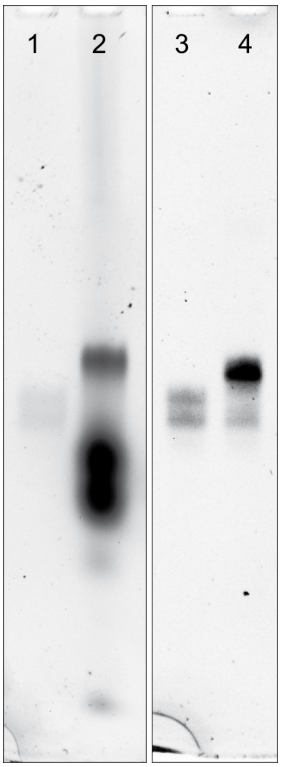


Figure 333. Full lanes (*SYBR Safe* stain) of the 20 % DPAGE analyses shown in Figure 70; Lane 1: purified **RNA^{CP}**; Lane 2: purified **RNA^{CP}+28**; Lane 3: purified **RNA^{CP}**; Lane 4: : purified **RNA^{CP}+57**.

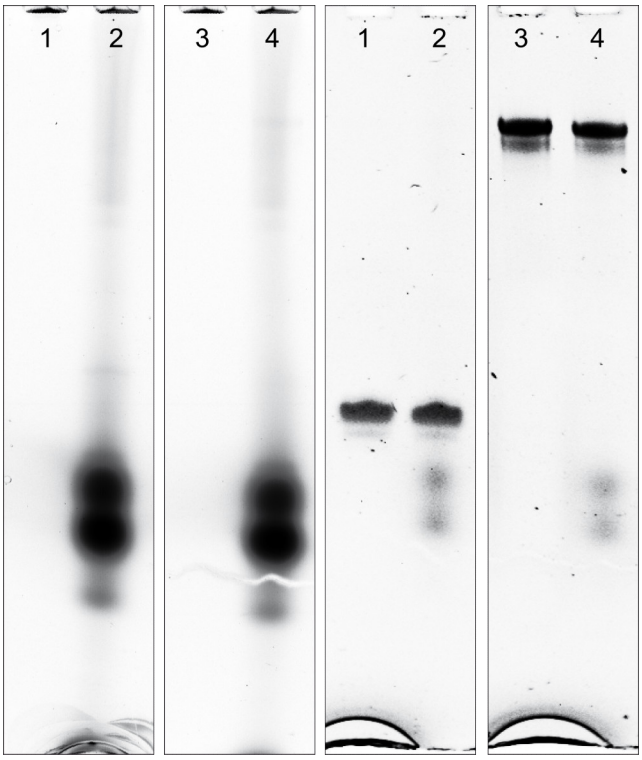


Figure 334. Full lanes of the fluorescence scan (left panels) and the *SYBR Safe* stain (right panels) of the 20 % DPAGE analysis shown in Figure 72; Lane 1: **RNA^{c}**, Lane 2: **RNA^{c}+28**; Lane 3: **tRNA^{c}**; Lane 4: **tRNA^{c}+28**.

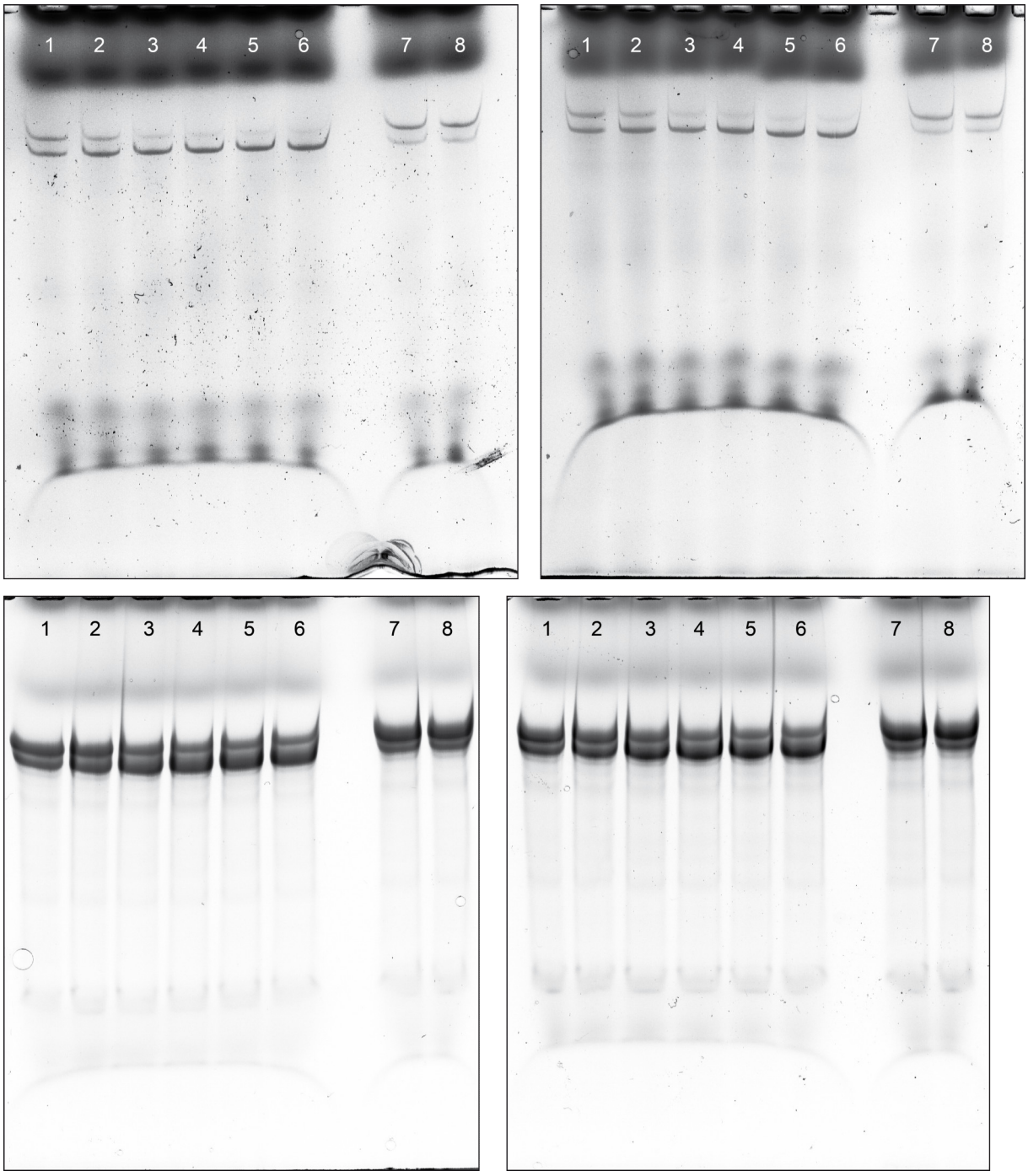


Figure 335. Full lanes (fluorescence scan) of 12 % DPAGE analyses for **glmS_RNA^{CP}-Tet-OG** cleavage data (1 of 3); **glmS_RNA^{CP}-Tet-OG** incubated with **1** after 1 min (Lane 1), 2 min (Lane 2), 5 min (Lane 3), 15 min (Lane 4), 30 min (Lane 5), and 60 min (Lane 6); Lane 7: **glmS_RNA^{CP}-Tet-OG** after 1 min; Lane 8: **glmS_RNA^{CP}-Tet-OG** after 60 min.

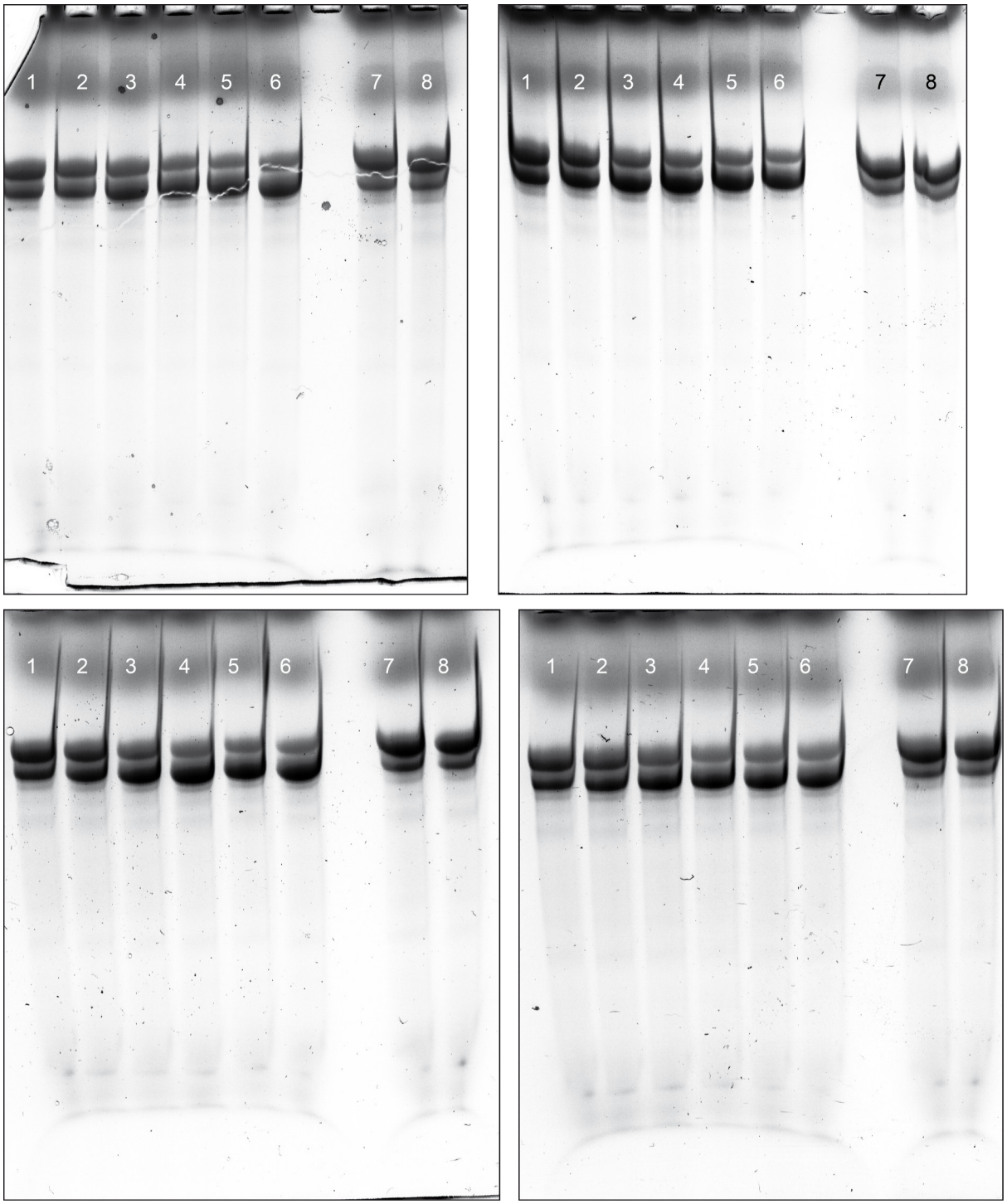


Figure 336. Full lanes (fluorescence scan) of 12 % DPAGE analyses for **glmS_RNA^{CP}-Tet-OG** cleavage data (2 of 3); **glmS_RNA^{CP}-Tet-OG** incubated with **1** after 1 min (Lane 1), 2 min (Lane 2), 5 min (Lane 3), 15 min (Lane 4), 30 min (Lane 5), and 60 min (Lane 6); Lane 7: **glmS_RNA^{CP}-Tet-OG** after 1 min; Lane 8: **glmS_RNA^{CP}-Tet-OG** after 60 min.

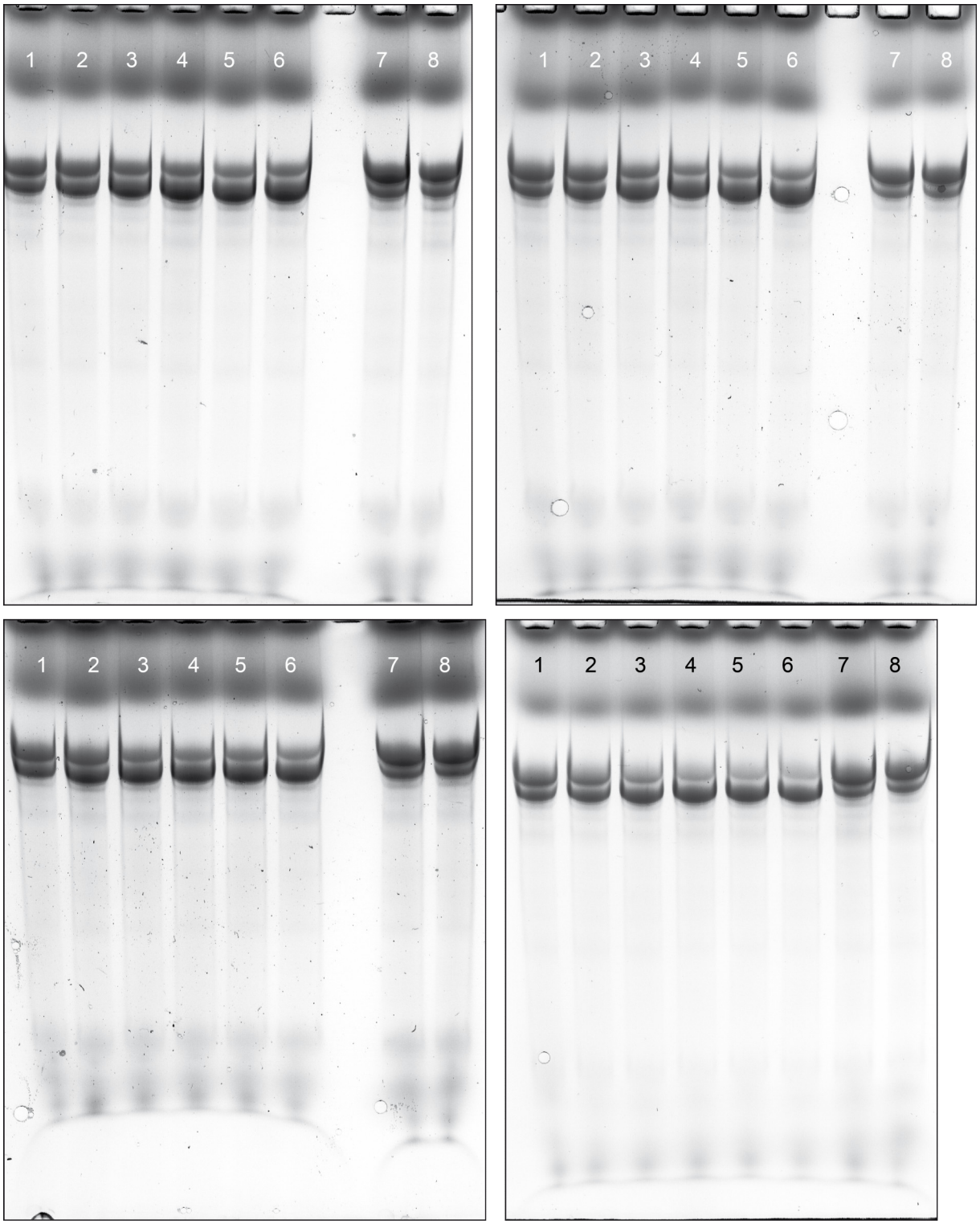


Figure 337. Full lanes (fluorescence scan) of 12 % DPAGE analyses for **glmS_RNA^{CP}-Tet-OG** cleavage data (3 of 3); **glmS_RNA^{CP}-Tet-OG** incubated with **1** after 1 min (Lane 1), 2 min (Lane 2), 5 min (Lane 3), 15 min (Lane 4), 30 min (Lane 5), and 60 min (Lane 6); Lane 7: **glmS_RNA^{CP}-Tet-OG** after 1 min; Lane 8: **glmS_RNA^{CP}-Tet-OG** after 60 min.

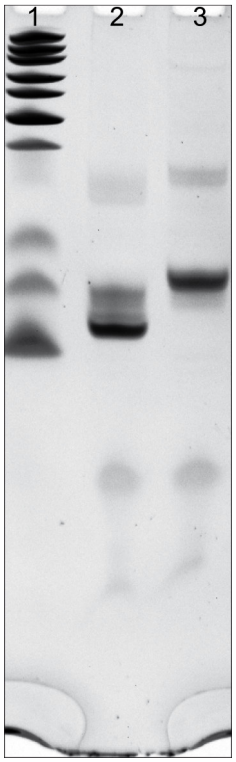


Figure 338. Full lanes of the 20 % DPAGE (*SYBR Safe* stain) as depicted in Figure 77; Lane 1: Marker; Lane 2: **RNA^{sc_n}**; Lane 3: **RNA^{sc_CP}-Tet-NO**.

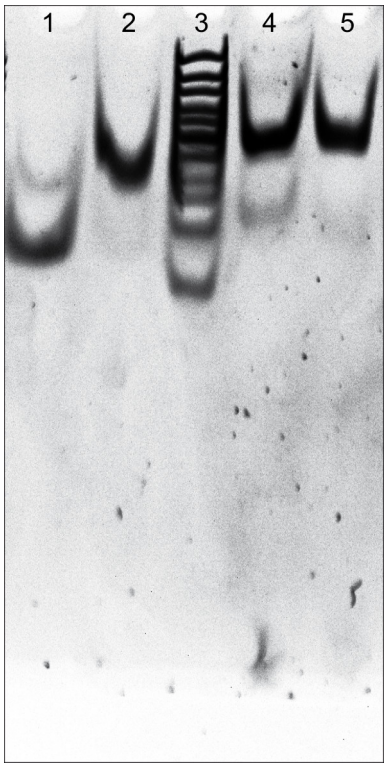


Figure 339. Full lanes of the 20 %native PAGE analysis (*SYBR Safe* stain) shown in Figure 82; Lane 1: **RNA^{sc_n}**; Lane 2: **RNA^{sc_NO}**; Lane 3: **RNA^{sc_ext_n}**; Lane 4: **RNA^{sc_ext_NO}**.

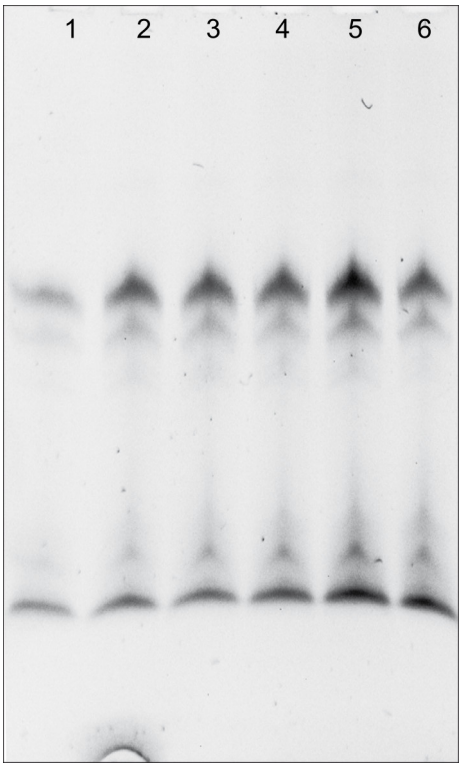


Figure 340. Complete lanes of 20 % DPAGE analysis (*SYBR Safe* stain) shown in Figure 84; Lane 1-6: **RNA^{sc_ext_NO}**.

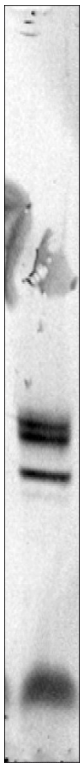


Figure 341. Full lane of the 20 % DPAGE analysis (*SYBR Safe* stain) as depicted in Figure 86.

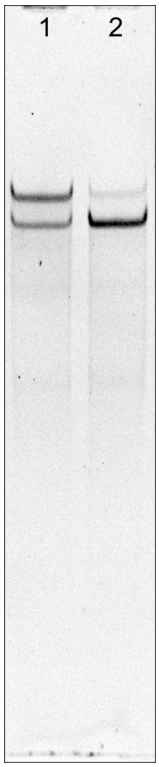


Figure 342. Full lanes of the 12 % DPAGE analysis (*SYBR Safe* stain) shown in Figure 91; Lane 1: **glmS_RNA^{NO}**; Lane 2: **glmS_RNA^{NO}**+200 µM GlcN6P (**1**).

Figure 343. Complete lanes of the 12 % DPAGE analysis (*SYBR Safe* stain) shown in Figure 92; Lane 1: **glmS_RNA^{n_short}**; Lane 2: **glmS_RNA^{NO_P1P4a_short}**; Lane 3: **glmS_RNA^{NO_P1P4L_short}**.

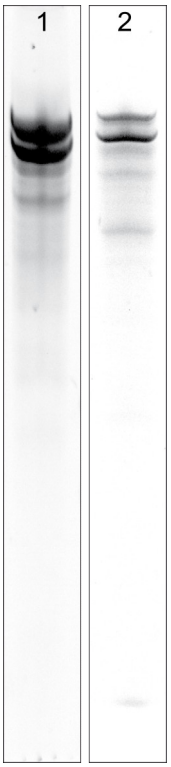


Figure 344. Full lanes of the 12 % DPAGE analyses (*SYBR Safe* stain) shown in Figure 95; Lane 1: **glmS_RNA^{NO_P1P4a}**; Lane 2: **glmS_RNA^{NO_P4RP4a}**.

Figure 345. Full lanes of the 12 % DPAGE analyses (*SYBR Safe* stain) shown in Figure 98; Lane 1: **glmS_RNA^{NO_P1P4a}**; Lane 2: **glmS_RNA^{NO_P4RP4a}**.

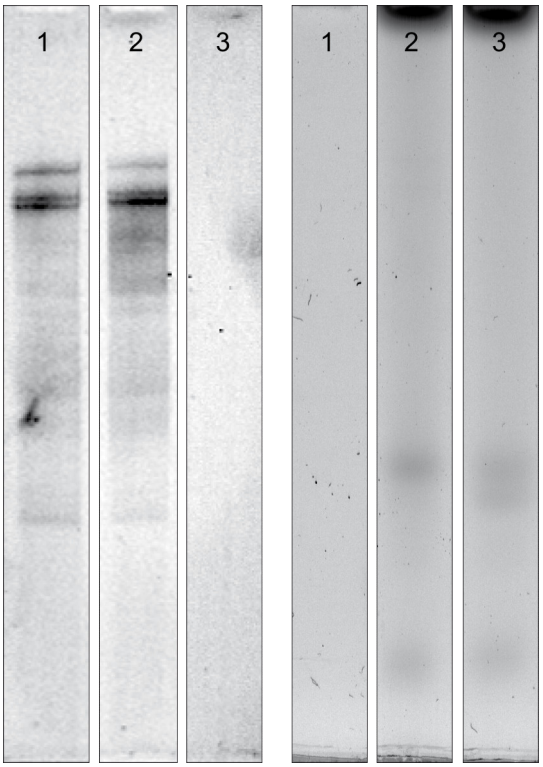


Figure 346. Complete lanes of 12 % DPAGE analysis (*SYBR Safe* stain left panels, fluorescence scan right panels) depicted in Figure 102; Lane 1: **glmS_RNA**$^{\text{CP\_P4RP4a\_HH}}$; Lane 2: **glmS_RNA**$^{\text{CP\_P4RP4a\_HH}}$**+59**; Lane 3: **59**.

Figure 347. Full lane of the 12 % DPAGE lane (*SYBR Safe* stain) depicted in Figure 103 showing **glmS_RNA**$^{\text{NO\_P4RP4a}}$.

Figure 348. Full lane of the 12 % DPAGE analysis (*SYBR Safe* stain) depicted in Figure 104 showing **glmS_RNA^{NO_P4a}**.

Agarose gels

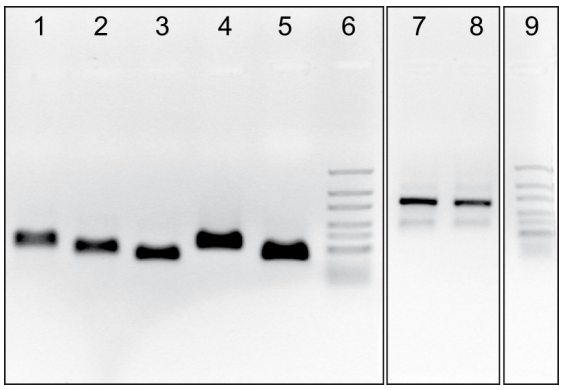


Figure 349. Full lanes of the 2 % agarose gels (EtBr stain) depicted in Figure 58; Lane 1: **glmS_DNA^{1n}**, Lane 2: **glmS_DNA^{2n}**; Lane 3: **glmS_DNA^{3n}**; Lane 4: **glmS_DNA^{1NaM}**; Lane 5: **glmS_DNA^{3NaM}**; Lane 6: Marker; Lane 7: Extension approach with unmodified templates and canonical rNTPs; Lane 8: Extension with d**NaM**-modified templates and d**TPT3** (**22**)/d**NaM** TP (**24**); Lane 9: Marker.

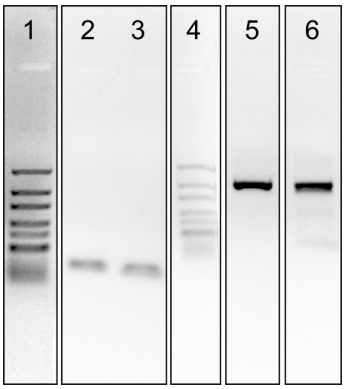


Figure 350. Full lanes of the 2 % agarose gel (EtBr stain) depicted in Figure 59; Lane 1: Marker; Lane 2: **Primer_glmS_FW**; Lane 3: **Primer_glmS_RV[1]**; Lane 4: Marker; Lane 5: **glmS_ DNA$^{n_full_sense/anti}$**; Lane 6: **glmS_DNA$^{full_sense/anti}$**.

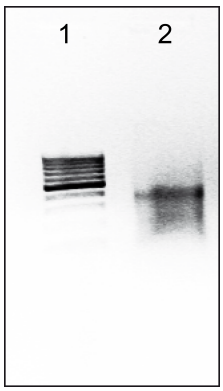


Figure 351. Full lanes of the 1 % agarose gel (EtBr stain) depicted in Figure 66; Lane 1: Marker; Lane 2: **Xist_RNA^{CP}**.

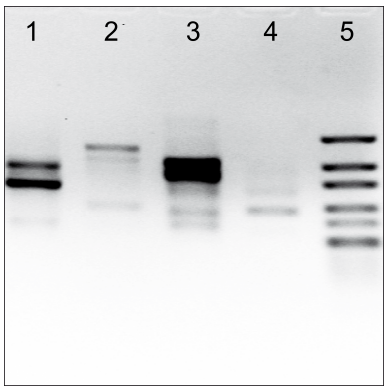


Figure 352. Full lanes of the 2 % agarose gel (EtBr stain) depicted in Figure 101; Lane 1: extension step towards **glmS_DNA^{n_P4RP4a_HH}**; Lane 2: amplification step towards **glmS_DNA^{n_P4RP4a_HH}**; Lane 3: extension step towards **glmS_DNA^{P4RP4a_HH}**; Lane 4: amplification step towards **glmS_DNA^{P4RP4a_HH}**; Lane 5: Marker.

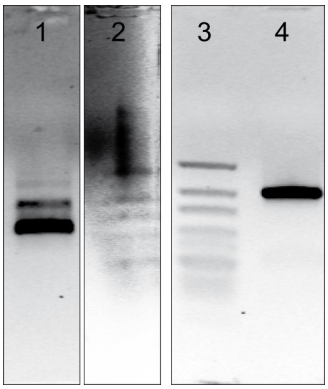


Figure 353. Complete lanes of the 2 % agarose analyses (EtBr stain) depicted in Figure 102; Lane 1: extension step towards **glmS_DNA^{P4RP4a_HH}**; Lane 2 and Lane 3: Marker; Lane 4: amplification step towards **glmS_DNA^{P4RP4a_HH}**.

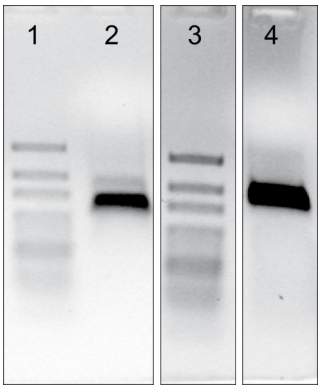


Figure 354. Complete lanes of the 2 % agarose analyses (EtBr stain) depicted in Figure 103; Lane 1: Marker; Lane 2: extension step towards **glmS_DNA^{P4RP4a}**; Lane 3: Marker; Lane 4: amplification step towards **glmS_DNA^{P4RP4a}**.

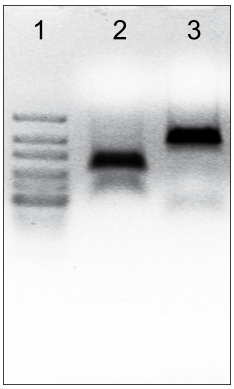
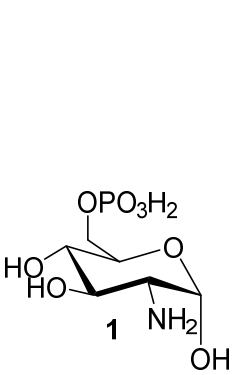


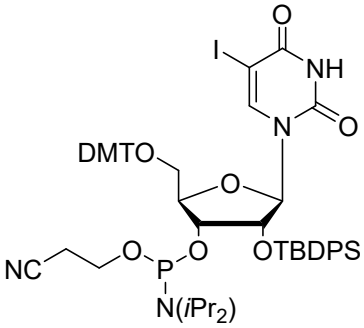
Figure 355. Full lanes of 2 % agarose analysis (EtBr stain) shown in Figure 104; Lane 1: Marker; Lane 2: extension step towards **glmS_DNA^{P4a}**; Lane 3: amplification step of **glmS_DNA^{P4a}**.

7.5 List of compounds

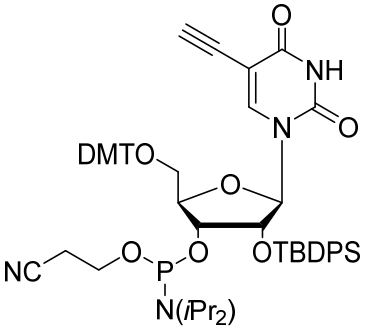
This section contains all compounds that have been mentioned with explicit chemical structure throughout this study.



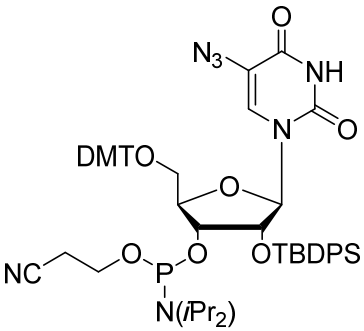
1



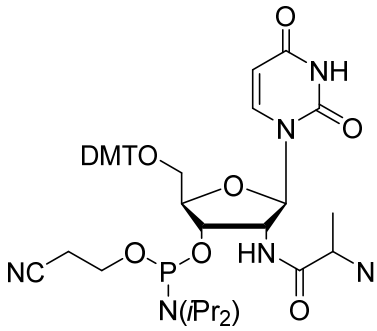
2



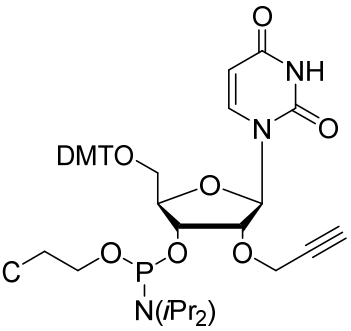
3



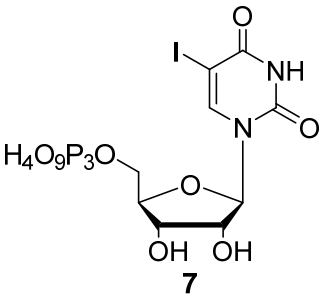
4



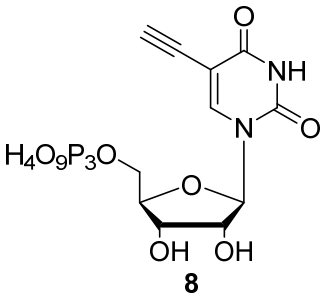
5



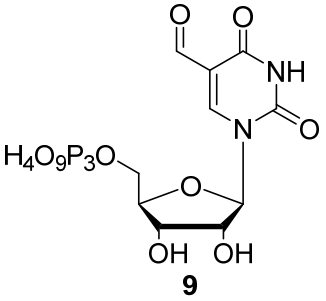
6



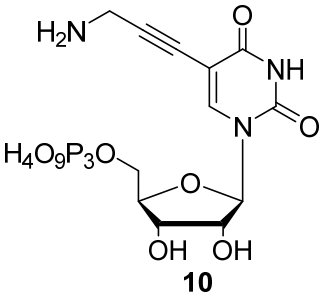
7



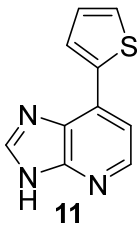
8



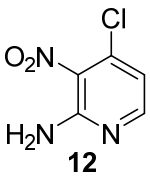
9



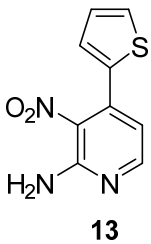
10



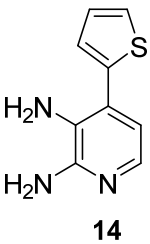
11



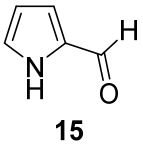
12



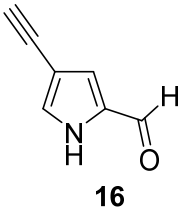
13



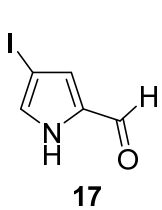
14



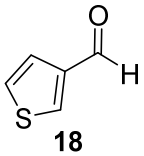
15



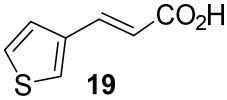
16



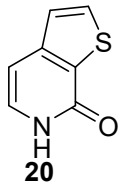
17



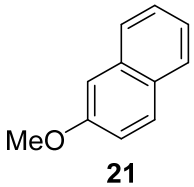
18



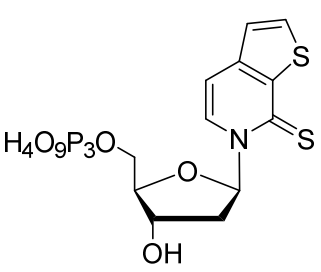
19



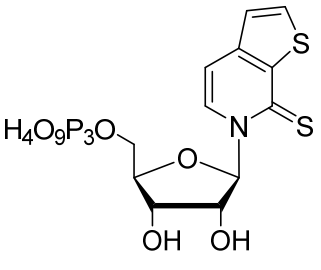
20



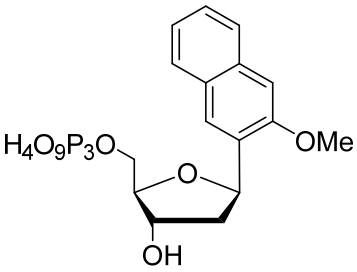
21



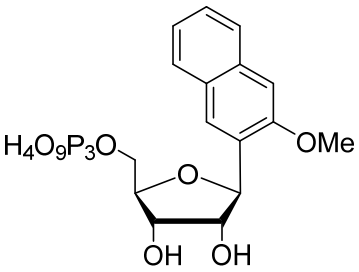
22



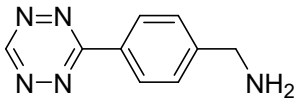
23



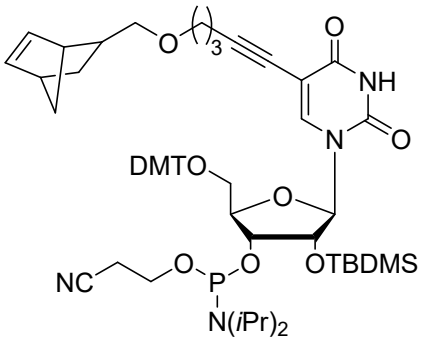
24



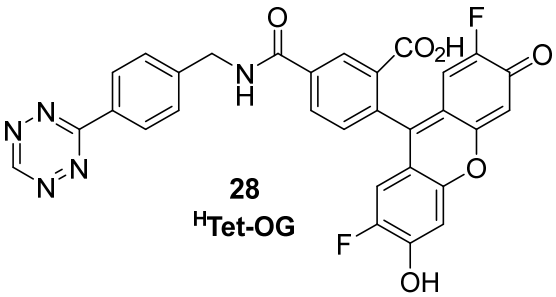
25



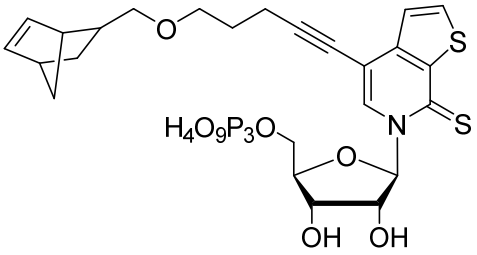
26



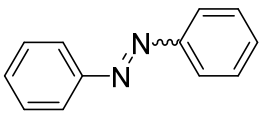
27
r**U**$^{\mathbf{NOR}}$ CEP



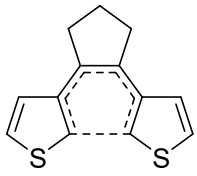
28
$^{\mathrm{H}}$**Tet-OG**



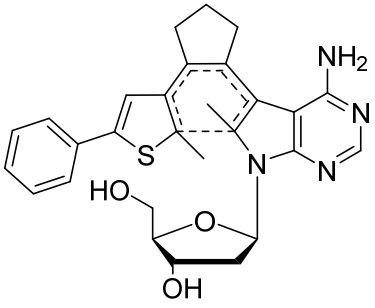
29
r**TPT3**$^{\mathbf{NOR}}$ TP



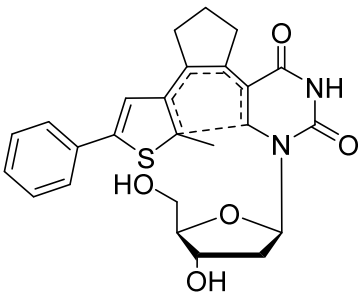
cis-/trans-**30**



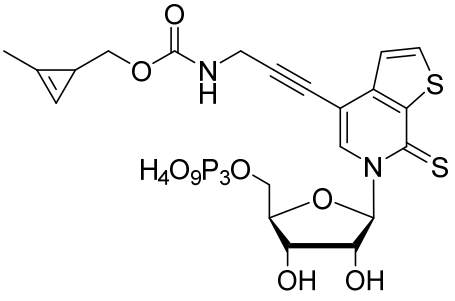
open-/*closed*-**31**



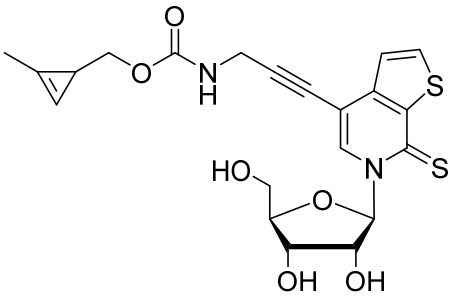
open-/*closed*-**32**



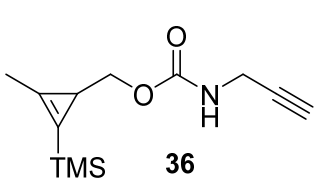
open-/*closed*-**33**



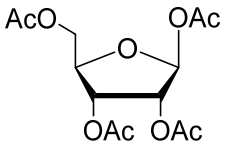
34
r**TPT3**$^{\mathbf{CP}}$ TP



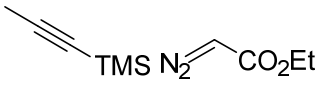
35
r**TPT3**$^{\mathbf{CP}}$



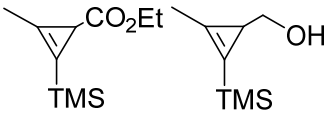
36



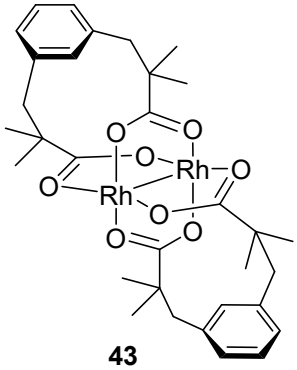
37



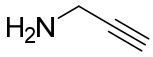
38 **39**



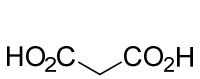
40 **41**



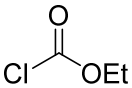
43



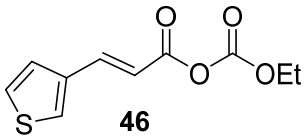
42



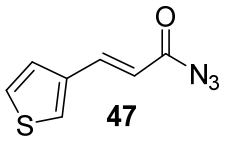
44



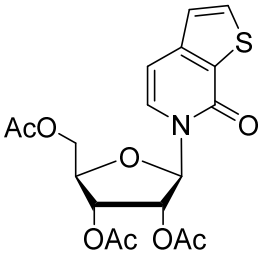
45



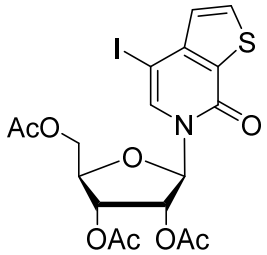
46



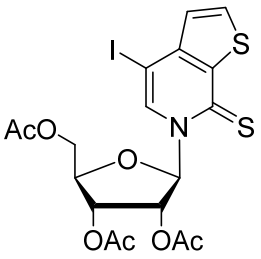
47



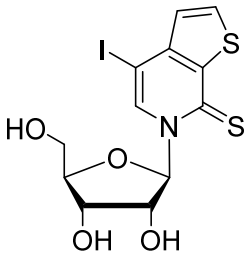
48



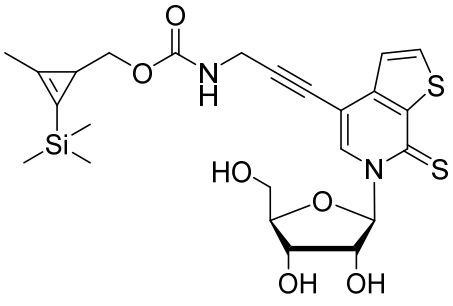
49



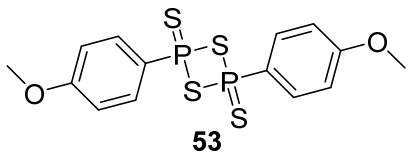
50



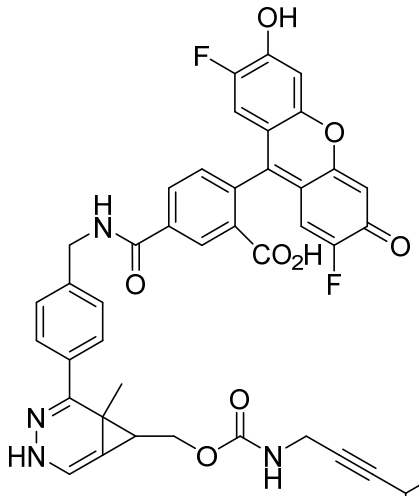
51



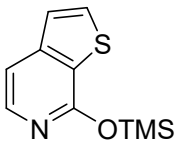
52



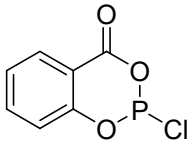
53



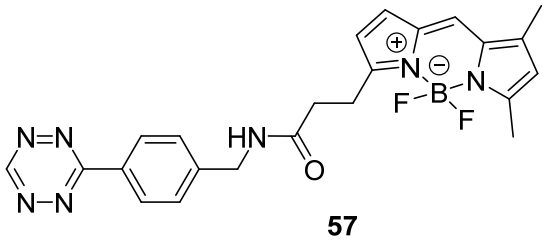
55



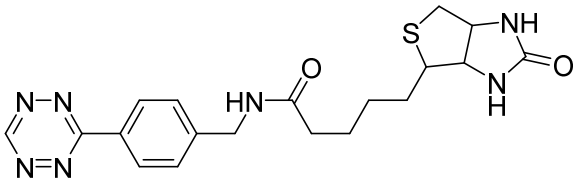
54



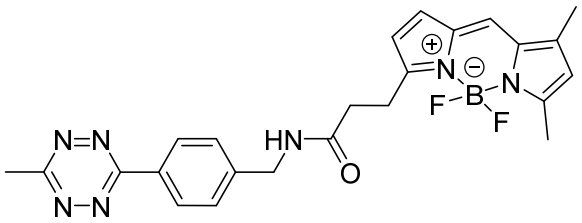
56



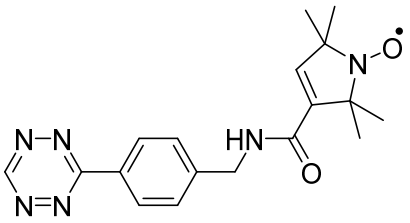
57
^{H}Tet-BDP



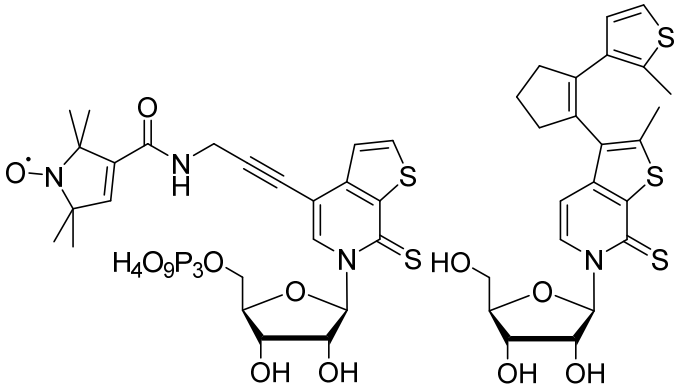
58



59
^{Me}Tet-BDP

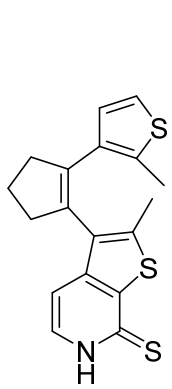


60
Tet^{NO}

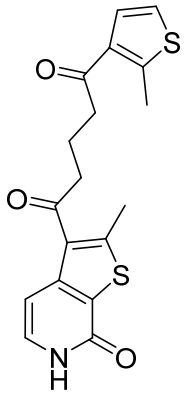


61
rTPT3^{NO} TP

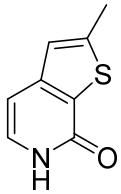
62
rTPT2^{PS}



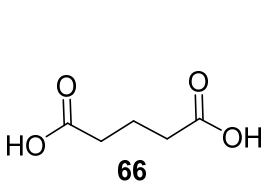
63



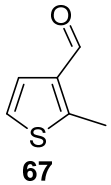
64



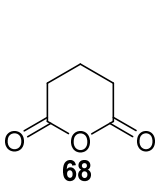
65



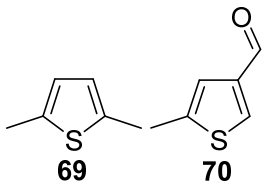
66



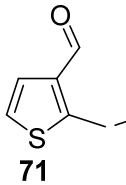
67



68

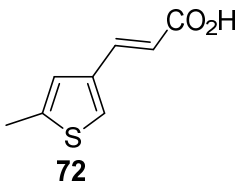


69

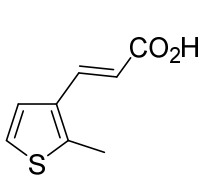


70

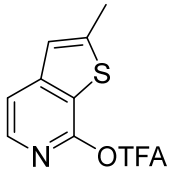
71



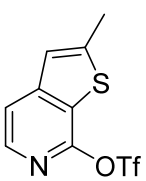
72



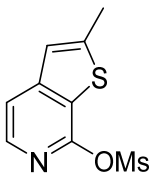
73



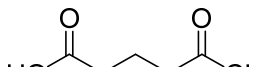
74



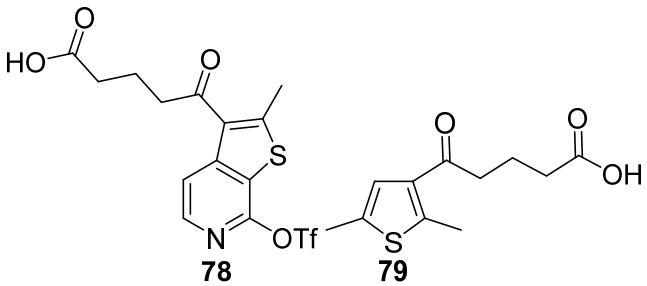
75



76

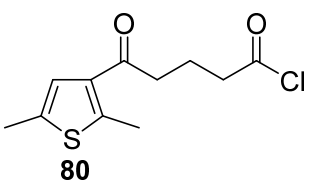


77

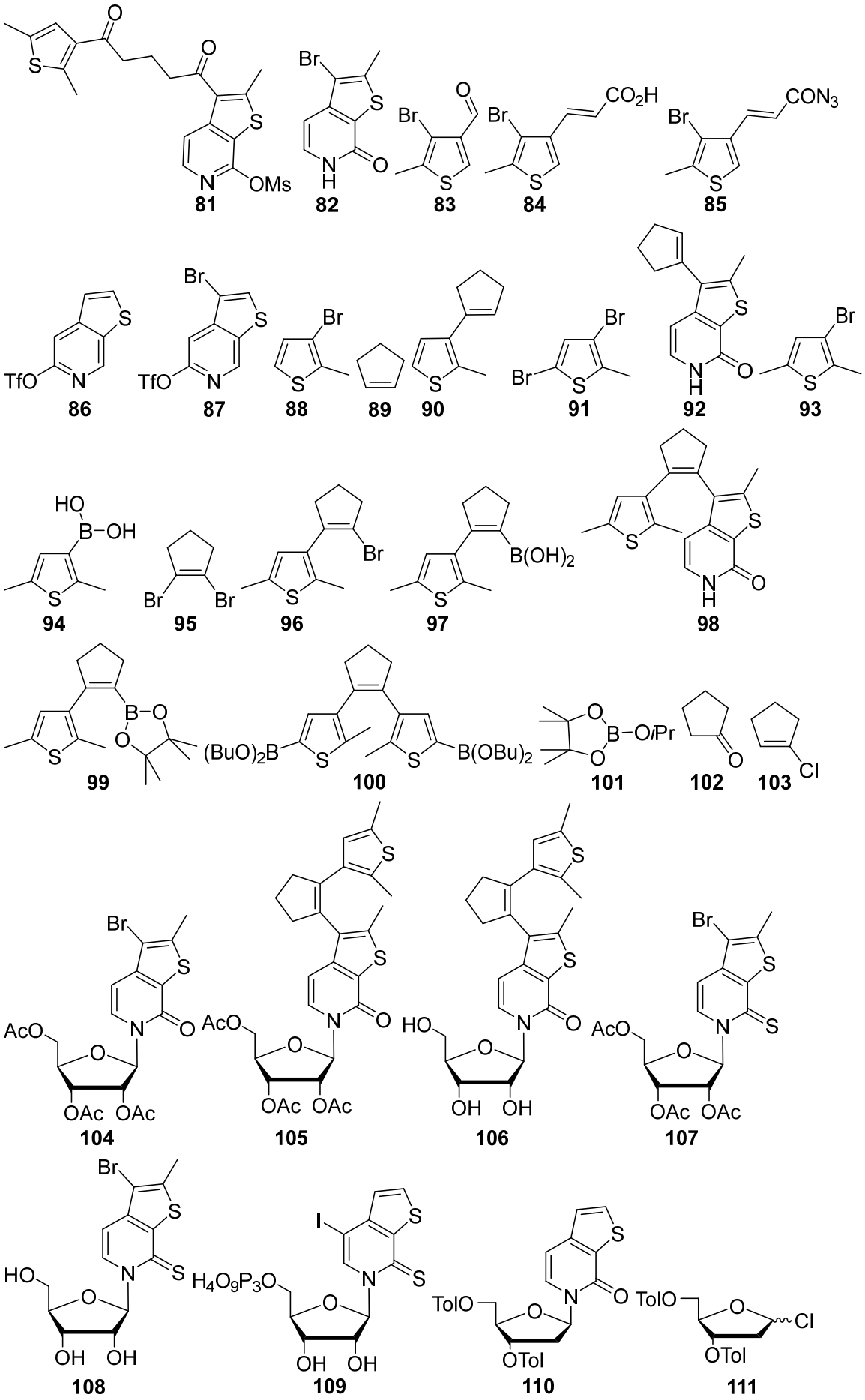


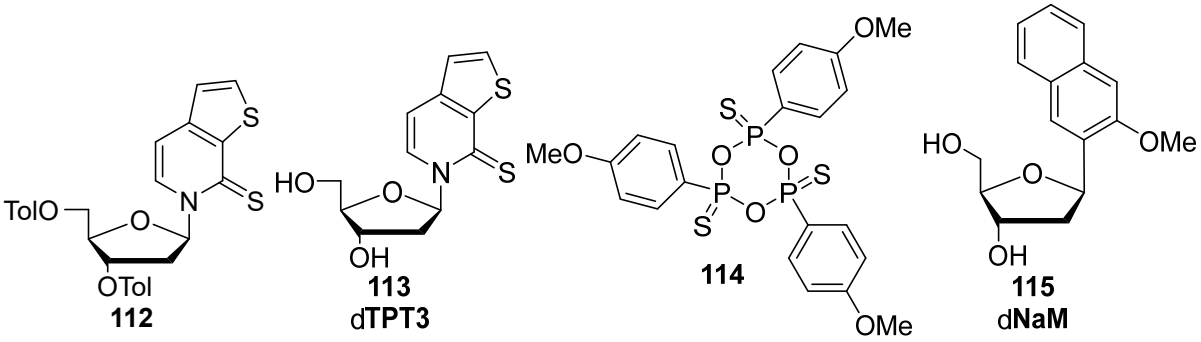
78

79

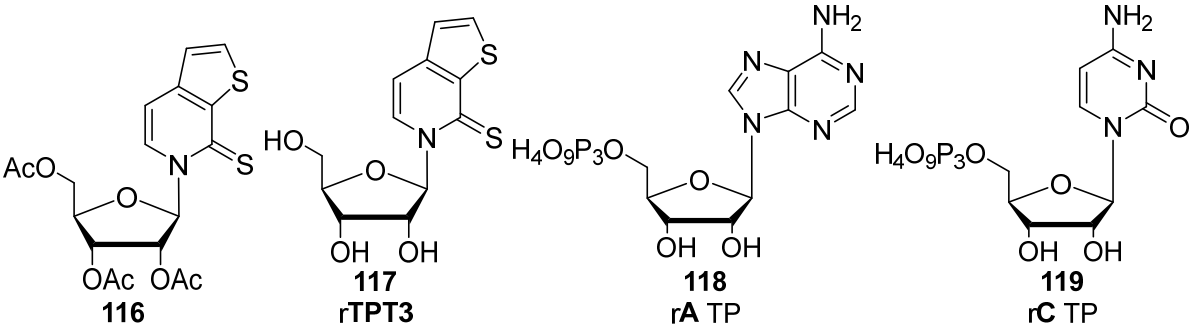


80

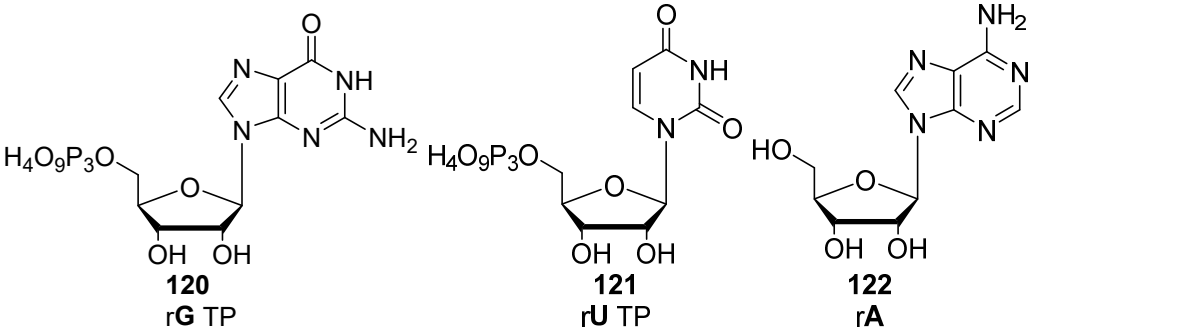

81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111



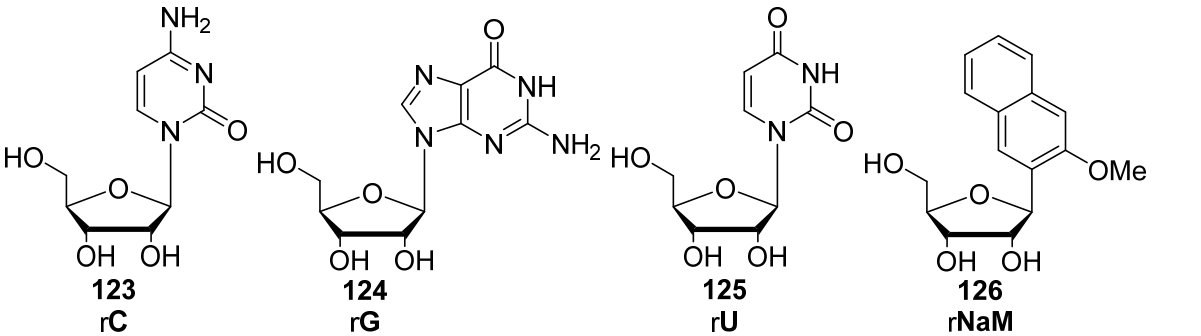
112 **113** d**TPT3** **114** **115** d**NaM**



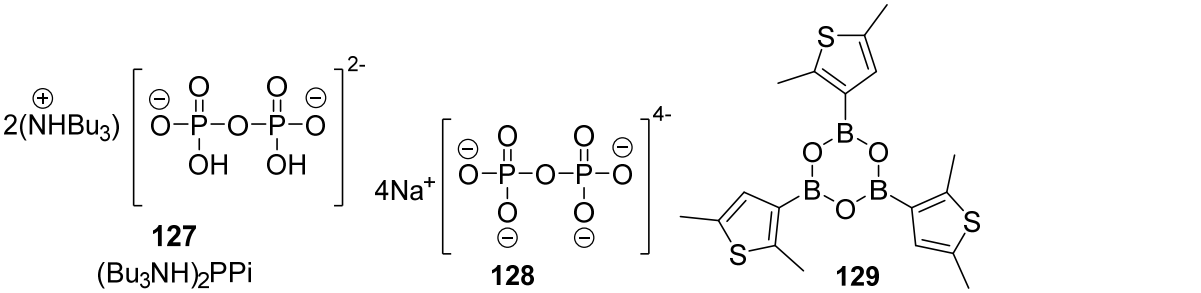
116 **117** r**TPT3** **118** r**A** TP **119** r**C** TP



120 r**G** TP **121** r**U** TP **122** r**A**



123 r**C** **124** r**G** **125** r**U** **126** r**NaM**



127 $(Bu_3NH)_2PPi$ **128** **129**

7.6 List of abbreviations

Table 33 shows all abbreviated and otherwise uncommon scientific terms used in this study.

Table 33. Abbreviations, acronyms, and terms taken from Latin language.

$(n\text{-}Bu_3NH)_2PPi$	tri-*n*-butylammonium pyrophosphate
3′-cP	2′,3′-cyclic phosphate
6-FAM	6-carboxyfluorescein
A	adenine
a priori	from the earlier
a.u.	arbitrary units
ab initio	from the beginning
AcOH	acetic acid
AEGIS	artificially expanded genetic information systems
AMBER	assisted model building with energy refinement
AMV	avian myeloblastosis virus
approx.	approximately
aq.	aqueous
$BF_3{\cdot}OEt_2$	boron trifluoride diethyletherate
BODIPY	boron-dipyrromethene
bp	base pair(s)
bpb	bromophenol blue
br s	broad singlet
BSA	*N*,*O*-bis(trimethylsilyl)acetamide
C	cytosine
cat.	catalyst or catalytic
CD	circular dichroism
CDI	carbonyldiimidazole
cDNA	complementary DNA
CEP	2-cyanoethyl-*N*,*N*-diisopropyl-phosphoramidite
CHARMM	chemistry at Harvard macromolecular mechanics
cw-EPR	continuous wave EPR
Cy	cyclohexane
d	doublet
d**A**	deoxyadenosine
DAPI	4′,6-diamidino-2-phenylindole
d**C**	deoxycytidine
ddH_2O	double distilled water

ddNTP	dideoxynucleoside triphosphate
DEAE	diethylaminoethyl
DFT	density functional theory
d**G**	deoxyguanosine
DIBALH	diisobutylaluminium hydride
DIPEA	diisopropylethyl amine
DMF	dimethylformamide
DMS	dimethyl sulfide
DMSO	dimethyl sulfoxide
DMT	dimethoxytrityl
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DP	diphosphate
DPAGE	denaturing polyacrylamide gel electrophoresis
ds	double stranded
d**T**	deoxythymidine
DTT	dithiothreitol
E. coli	*Escherichia coli*
e.g.	for example (Lat. *exempli gratia*)
EDA	ethyldiazo acetate
EDTA	ethylenediaminetetraacetic acid
EI	electron ionization
Electrophoretic mobility shift assay	EMSA
EPR	electron paramagnetic resonance
eq.	equivalent(s)
ESI	electron spray ionization
et al.	and others (Lat. *et alii*)
Et_2O	diethyl ether
EtBr	ethidium bromide
EtOAc	ethyl acetate
EtOH	ethanol
EWG	electron withdrawing group
ExSELEX	genetic alphabet expansion SELEX
FA	formamide
FISH	fluorescence *in situ* hybridization

FRET	Förster resonance energy transfer
G	guanine
GFP	green fluorescent protein
GlcN6P	Glucosamine-6-phosphate
glmS	glucosamine-6-phosphate synthase
GNA	glycine nucleic acid
GRMOACS	Groningen machine for chemical simulations
H-bonding	hydrogen bonding
HDV	hepatitis delta virus
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HH	hammerhead
HIV	human immunodeficiency virus
HMDS	hexymethyldisilazane
HMQC	heteronuclear multiple quantum coherence
HPLC	high performance liquid chromatography
HR	high resolution
HSQC	heteronuclear single quantum coherence
Hz	Hertz
i.e.	that is (Lat. *id est*)
IEDDA	inverse electron-demand Diels-Alder
in situ	on site
in vitro	in the test tube
in vivo	in the living
iPP	inorganic pyrophosphatase
iso-C	isocytidine
iso-G	isoguanosine
J	coupling constant
kb	kilo base(s)
lncRNA	long non-coding RNA
M	Marker
$M_{calcd.}$	calculated mass
MD	molecular dynamics
Me_3PO_4	trimethyl phosphate
MeCN	acetonitrile
MeI	methyl iodide
MeOH	methanol
miRNA	micro ribonucleic acid

MMLV	murine leukemia virus
mod	modified
MP	monophosphate
mRNA	messenger ribonucleic acid
MS	mass spectrometry
Ms	Methanesulfonatem mesylate
MWD	multi wavelength detector
n.d.	not determined/determinable
NaOAc	sodium acetate
NaOMe	sodium methoxide
NBS	*N*-bromosuccinimide
NBu_3	tributylamine
n-BuLi	*n*-butyllithium
ncRNA	non-coding RNA
NEt_3	triethylamine
NH_4OAc	ammonium acetate
NHS	*N*-hydroxysuccinimide
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser spectroscopy
nt	nucleotide(s)
o/n	overnight
OMe	methoxy
p	pentet
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PELDOR	pulsed electron-electron double resonance
Ph	phenyl
Ph_2O	diphenyl ether
PNA	peptide nucleic acid
Pol	polymerase
PP_i	inorganic pyrophosphate
ppm	parts per milion
py	pyridine
q	quintet
quant.	quantitative
r.t.	room temperature

r**A**	adenosine
r**C**	cytidine
rel.	relative
R_f	retention factor
r**G**	guanosine
$Rh_2(esp)_2$	bis[rhodium(α,α,α′,α′-tetramethyl-1,3-benzene-dipropionic acid)]
$Rh_2(OAc)_4$	rhodium(II) acetate dimer
RNA	ribonucleic acid
RNase	ribonuclease
RNasin	ribonuclease inhibitor
rNTP	ribonucleoside triphosphate
rRNA	ribosomal RNA
RT	reverse transcriptase
rtc	reverse transcription
r**U**	uridine
rz	ribozyme
s	singlet
SAR	structure activity relationship
sat.	saturated
s-BuLi	*sec*-butyllithium
SD	standard deviation
SELEX	systematic enrichment of ligands by exponential enrichment
siRNA	small interfering ribonucleic acid
soln.	solution
ss	single stranded
SS II/IV	Superscript II/IV
SSO	semi-synthetic organism
T	thymine
t	triplet
Taq	Thermus aquaticus
TB	tris-borate
TBAF	tetra-*n*-butylammonium fluoride
TBDMS	*tert*-butyldimethylsilyl
TBDPS	*tert*-butyldiphenylsilyl
TBE	tris-borate-EDTA

*t*BuOK	potassium *tert*-butoxide
TEAB	triethylammonium bicarbonate
TEMED	tetramethylethylenediamine
TEMPYO	2,2,5,5-tetramethyl-3-carboxylpyrrolin-1-oxyl
Tf	triflate
TFA_2O	trifluoroacetic acid anhydride
TFMSA	trifluormethanesulfonic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMEDA	*N*,*N'*,*N''*,*N'''*-tetramethylethylendiamine
TMS	trimethylsilyl
TNA	threose nucleic acid
Tol	toluoyl
TP	triphosphate
TPPTS	3,3',3''-phosphanetriyltris(benzenesulfonic acid)
t_R	retention time
tRNA	transfer ribonucleic acid
U	uracil
UB(P)s	unnatural base (pair)s
unmod	unmodified
UTR	untranslated region
UV	ultraviolet
v/*v*	volume by volume
V2G	Volcano 2G
VEGF	vascular endothelial growth factor
VIS	visible light
w/	with
w/*o*	without
w/*v*	weight by volume
xc	xylene cyanol

7.7 List of tables

7.8 List of figures

8 Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit sowohl ohne unerlaubte Hilfe Dritter als auch ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche kenntlich gemacht.

Bonn, den 05.09.2019

_____

Frank Eggert